Probabilistic Modeling of Landmark Distances and Structure for Anomaly-proof Landmark Detection

Shouhei Hanaoka[1], Yoshitaka Masutani[1,2], Mitsutaka Nemoto[1], Yukihiro Nomura[1], Takeharu Yoshikawa[3], Naoto Hayashi[3], Naoki Yoshioka[4], Kuni Ohtomo[1,2]

[1] Department of Radiology, [2] Division of Radiology and Biomedical Engineering, Graduate School of Medicine, [3] Department of Computational Diagnostic Radiology and Preventive Medicine and [4] Department of Integrated Imaging Informatics, The University of Tokyo hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
hanaoka-tky@umin.ac.jp

Abstract. A combinatorial optimization algorithm for detecting multiple anatomical landmarks is presented. It can determine the positions of over 100 landmarks concurrently, taking spatial correlations of all landmark pairs into account. Provided that a set of landmark candidate lists is given by sensitivity-optimized single-landmark detectors, the proposed algorithm can find the most probable combination of them through solving a MAP estimation-based combinatorial optimization problem. Additionally, it is designed to handle subjects with "segmentation anomaly of the spinal column," a common anatomical anomaly of the spine. The proposed system was evaluated with 156 landmarks in 50 datasets, using virtually created detector output sets. In the result, the algorithm achieved 97.6% of spinal anomaly estimation accuracy even with 50 points of candidates given per landmark, as well as 96.2% of accuracy in landmark candidate selection. From these results, usefulness of the proposed algorithm for subjects with spinal anomaly was suggested.

Keywords: Landmark, Combinatorial optimization, MAP estimation, Anatomical anomaly, Computed tomography, Spine

1 Introduction

Landmark point detection algorithms are extensively researched and widely used in various medical image processing applications. However, it is a difficult task to detect a large number of landmarks correctly, because the human body includes a lot of similar regions sharing their appearances. Even most of frequently-used landmarks do not have truly unique local shape or intensity. Consequently, detection results often include a certain number of false positive candidates.

Furthermore, some sort of important anatomical landmarks can intrinsically be non-existent, not only in patients due to pathological condition, but even in healthy subjects due to anatomical anomalies (fig. 1a). Confirming such a situation is very difficult by detecting each landmark independently and sequentially.

One solution is to determine the entire landmark positions simultaneously, taking their spatial relation into account.   It can be done by dividing the whole problem into 2 sequential phases: the individual landmark detection phase and the combinatorial optimization phase (figs. 1b, 1c).   In the former phase, each landmark is detected by a single detector.   Each detector is optimized to maximize the sensitivity, not the specificity, so the detection result forms a candidate list which includes a lot of false positives.   In the latter phase, the best combination of choice from all candidate lists is selected with the use of *a priori* knowledge on the inter-landmark relationship. Defect of any landmark can also be detected through this optimization phase.
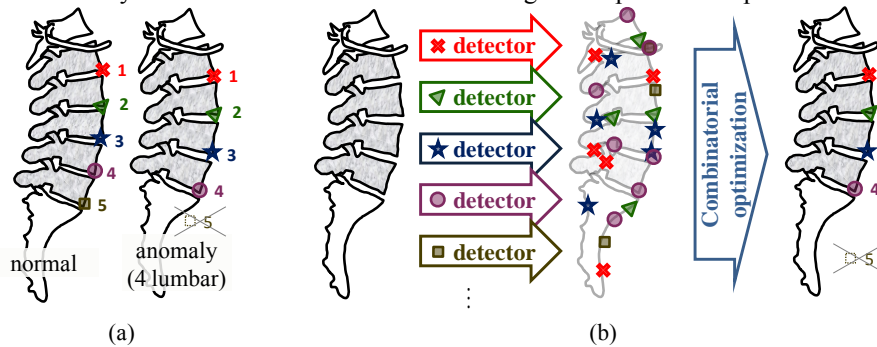
Fig. 1. (a) An example of anatomical anomaly and landmark deficit in the 5th lumbar vertebra. (b) A schema of the framework of the proposed method. Note that, for the non-existent landmark (□), all detected candidates are rejected by the following combinatorial optimization.

Among related studies, Seifert et al. reported a framework to automatically detect 19 landmarks. [1]   In their approach, a belief propagation algorithm is used with prior knowledge about landmarks' geometric relationships such as "to the right of", "close to", etc.   Though their result was excellent, the prior knowledge used and the way to build it were not described precisely.

In this paper, we introduce a framework to determine over 100 landmark positions reliably.   Especially, we focus upon our method to solve the landmark-set combinatorial optimization problem with use of maximum *a posteriori* (MAP) estimation.   The method was evaluated with 156 landmarks in 50 human body CT image datasets.   Virtually-created detector outputs were used in this study, rather than those of real detectors [2], in order to evaluate the ability of our framework to handle a large number of false candidates.   Furthermore, handling of common anomaly (alteration in the number of thoracic/lumbar vertebrae) was also evaluated.

2   Methods

2.1. Definition

2.1.1. Detector output.   Firstly, we defined a landmark detector generally.   In our framework, each detector has to output not only a series of candidate positions, but also estimated probabilities of them.   Based on this, outputs from one detector (for the m-th landmark) are defined as:

N_m	number of the candidates	$(0 \leq N_m)$
$\mathbf{c}_m^{\ i}$	coordinates of the i-th candidate	$(1 \leq i \leq N_m)$
$p_m^{\ i}$	detector-estimated probability of the i-th candidate	$(0 < p_m^{\ i} \leq 1)$
$p_m^{\ \Im}$	probability that "none of candidates $\mathbf{c}_m^{\ i}$ is correct"	$(0 \leq p_m^{\ \Im} \leq 1)$

The whole set of output is also defined as $I_m = \left\{ N_m, \mathbf{c}_m^{\ 1}, \ldots, \mathbf{c}_m^{\ N_m}, p_m^{\ 1}, \ldots, p_m^{\ N_m}, p_m^{\ \Im} \right\}$. Any output I_m must satisfy $p_m^{\ \Im} + \sum p_m^{\ i} = 1$.

Letting the true landmark position be $\mathbf{x}_m$, the detector-estimated probability $p_m^{\ i}$ can be interpreted as a conditional probability with a certain detector output I_m.

$$p_m^{\ i} = p\left(\mathbf{x}_m = \mathbf{c}_m^{\ i} \mid I_m \right) \tag{1}$$

If $p_m^{\ \Im} > 0$, it indicates that there are some possibilities of "no true landmark position is included in the candidate set $\left\{ \mathbf{c}_m^{\ 1}, \ldots, \mathbf{c}_m^{\ N_m} \right\}$."   The proposed algorithm can consider such a situation, and handles it as one extra state which will be represented as "$\mathbf{x}_m = \Im$" in this paper.   The state $\Im$ can be a "true answer" when the target landmark is out of the imaging range, or corresponding landmark does not exist anywhere (e.g., due to some anatomical anomaly or pathological condition).   In this describing method, the probability $p_m^{\ \Im}$ can also be interpreted as a conditional probability as

$$p_m^{\ \Im} = p\left(\mathbf{x}_m = \Im \mid I_m \right) \tag{2}$$

2.1.2. The prior probability distribution of LM positions.   Secondly, the prior probability function for all possible landmark position sets has to be defined in advance.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ be the positional vectors of total M landmarks, and $\mathbf{X} = (\mathbf{x}_1^{\ t} \, \mathbf{x}_2^{\ t} \ldots \mathbf{x}_M^{\ t})^t$ be the concatenated form of them.   Each element $\mathbf{x}_m$ can be regarded as a stochastic variable whose domain is $\mathbf{x}_m \in \left\{ \Im, \mathbf{c}_m^{\ 1}, \ldots, \mathbf{c}_m^{\ N_m} \right\}$.   The aim is to approximate the prior probability distribution of $\mathbf{X}$ as a single function $p(\mathbf{X})$, in order to use in the following MAP estimation.

In this study $p(\mathbf{X})$ is defined as a function of the squared distances between all landmark pairs.   When any $\mathbf{x}_m$ ($1 \leq m \leq M$) satisfies $\mathbf{x}_m = \Im$, however, the corresponding term of $p(\mathbf{X})$ is replaced by a conventionally-defined constant term.   The calculation method of $p(\mathbf{X})$ without considering $\Im$ is discussed in the next chapter, which will be followed by the general definition of $p(\mathbf{X})$.

(i) *When* $\mathbf{x}_m \neq \mathfrak{I},\ \forall m.$ Let

$$d_{i,j} = \left| \mathbf{x}_i - \mathbf{x}_j \right|^2 \quad (1 \leq i < j \leq M) \tag{3}$$

be the squared distance between the *i*-th and *j*-th landmarks.   Note that $d_{i,j}$ is defined only if both $\mathbf{x}_i$ and $\mathbf{x}_j$ are not $\mathfrak{I}$.   Then, the distance is normalized by its average $\mathrm{E}(d_{i,j})$ and variance $\mathrm{V}(d_{i,j})$ in the training datasets.   The normalized distance $g_{i,j}$ follows the equation

$$g_{i,j} = \frac{d_{i,j} - \mathrm{E}(d_{i,j})}{\sqrt{\mathrm{V}(d_{i,j})}} \qquad \text{if } \mathbf{x}_i \neq \mathfrak{I}, \mathbf{x}_j \neq \mathfrak{I}. \qquad (1 \leq i < j \leq M) \tag{4}$$

Let a vector $\mathbf{G} = (g_{1,2}\ g_{1,3} \ldots\ g_{i,j} \ldots g_{M-1,M})^t$ be the concatenated normalized squared distances between all landmark pairs.   Note that the vector $\mathbf{G}$ has $_M\mathrm{C}_2 = \dfrac{M(M-1)}{2}$ of elements.   Then, the prior probability distribution $p(\mathbf{X})$ was approximated by a multivariate normal distribution of the vector $\mathbf{G}$.   That is,

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2} \mathbf{G}^t \mathbf{V}^{-1} \mathbf{G} \right) \tag{5}$$

where $\mathbf{V}$ represents the covariance matrix of $\mathbf{G}$, which is calculated from training sets, and $|\mathbf{V}|$ is the determinant of $\mathbf{V}$.   The size of the matrix $\mathbf{V}$ is $_M\mathrm{C}_2 \times {}_M\mathrm{C}_2$.

(ii) *General definition*.   Eq. (5) can be written in an extended form

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} g_{i,j} \cdot \left\{ \mathbf{V}^{-1} \right\}_{(i,j),(k,l)} \cdot g_{k,l} \right) \tag{6}$$

where $\left\{ \mathbf{V}^{-1} \right\}_{(i,j),(k,l)}$ is the corresponding element of the matrix $\mathbf{V}^{-1}$.   Here, the summed term $g_{i,j} \cdot \left\{ \mathbf{V}^{-1} \right\}_{(i,j),(k,l)} \cdot g_{k,l}$ is not available when any of $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$ or $\mathbf{x}_l$ is $\mathfrak{I}$.   In order to define $p(\mathbf{X})$ generally, we replaced them as follows:

$$p(\mathbf{X}) = \frac{1}{(\sqrt{2\pi})^{_M\mathrm{C}_2} \cdot \sqrt{|\mathbf{V}|}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \sum_{k=1}^{M-1} \sum_{l=k+1}^{M} t_{(i,j),(k,l)} \right) \tag{7}$$

$$t_{(i,j),(k,l)} = \begin{cases} g_{i,j} \cdot \left\{ \mathbf{V}^{-1} \right\}_{(i,j),(k,l)} \cdot g_{k,l} & \text{if all of } \mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \text{ and } \mathbf{x}_l \text{ are not } \mathfrak{I} \\ 0 & \text{otherwise, and } (i,j) \neq (k,l) \\ \gamma & \text{otherwise, and } (i,j) = (k,l) \end{cases}$$

When $(i,j) \neq (k,l)$, the summed term $t_{(i,j),(k,l)}$ evaluates how $g_{i,j}$ and $g_{k,l}$ are correlated as like that of training sets.   When no prior information on $g_{i,j}$ (or $g_{k,l}$) is available, the expected value of this term should be zero.   That is why it should be replaced by zero in case of $\mathfrak{I}$.

On the other hand, when $(i,j) = (k,l)$, the term $t_{(i,j),(i,j)} = \left\{ \mathbf{V}^{-1} \right\}_{(i,j),(i,j)} \cdot g_{i,j}^2$ always has a positive value which evaluates how the distance $d_{i,j}$ varies from that of the training sets.   Therefore, we replaced it by a positive constant γ.   Because increasing this

parameter will reduce the probability $p(\mathbf{X})$ for any $\mathbf{x}_m = \Im$, γ can be regarded as a parameter which controls how the algorithm avoids the state $\Im$, or how it prefers any detected candidate $\mathbf{c}_m^i$, for each landmark.   We empirically selected $\gamma=2$ based on our preliminary experiments (the data is not shown).

2.1.3. *a posteriori* probability distribution.   Once the detector outputs are given, the probability *a posteriori* can be calculated with Bayes' theorem.   From the series of detector outputs I_1, I_2, …, I_M and the prior probability distribution $p(\mathbf{X})$, the posterior probability can be calculated by the Bayes' as:

$$p(\mathbf{X} \mid I_1, I_2, I_3,..., I_M) \;=\; \frac{p(I_1, I_2, I_3,..., I_M \mid \mathbf{X}) \cdot p(\mathbf{X})}{p(I_1, I_2, I_3,..., I_M)} \tag{8}$$

The denominator is constant and independent of $\mathbf{X}$.   Therefore, the maximum *a posteriori* estimation of landmark position set $\mathbf{X}$ is as follows:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X} \mid I_1, I_2, I_3,..., I_M) = \arg\max_{\mathbf{X}} p(I_1, I_2, I_3,..., I_M \mid \mathbf{X}) \cdot p(\mathbf{X}) \tag{9}$$

We assumed that all detector outputs are independent of each other, as well as independent on the positions of the other landmarks.   It means that each detector output is only dependent on the corresponding landmark position.   Then the term $p(I_1, I_2, I_3,... I_M \mid \mathbf{X})$ in (9) can be divided into the product of single-landmark conditional probabilities $p(I_m \mid \mathbf{x}_m)$.   That is,

$$p(I_1, I_2, I_3,..., I_M \mid \mathbf{X}) \;=\; \prod_{m=1}^{M} p(I_m \mid \mathbf{x}_m) = \prod_{m=1}^{M} \frac{p(\mathbf{x}_m \mid I_m) \cdot p(I_m)}{p(\mathbf{x}_m)} \tag{10}$$

The term $p(I_m)$ in this formula is independent of $\mathbf{X}$, so it can be ignored through the MAP estimation.   The denominator $p(\mathbf{x}_m)$ is also can be ignored, because the term is the probability distribution of one single landmark position $\mathbf{x}_m$ *without* any prior information.   Ignoring the term is equivalent to regarding it to be constant and homogeneous anywhere.   The remaining term $p(\mathbf{x}_m \mid I_m)$ is the detector-estimated probability as described in Eqs. (1) and (2).   That is,

$$p(\mathbf{x}_m \mid I_m) = \begin{cases} p_m^{\,i} & \text{if } \mathbf{x}_m = \mathbf{c}_m^{\,i}, \exists i \\ p_m^{\,\Im} & \text{if } \mathbf{x}_m = \Im \end{cases} \tag{11}$$

In conclusion, the MAP estimation will be performed by the following formula:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{X} \mid I_1, I_2, I_3,..., I_M) = \arg\max_{\mathbf{X}} \left[\left\{ \prod_{m=1}^{M} p(\mathbf{x}_m \mid I_m) \right\} \cdot p(\mathbf{X}) \right] \tag{12}$$

2.2. Implementation

2.2.1. Tikhonov's regularization.   In practice, the dimension of **G** in Eq. (5) (or Eq. (7)), which equals $_MC_2 = \frac{M(M-1)}{2}$, can be much greater than the number of training cases N.   In such a case, the covariance matrix **V** will be rank-deficit and has no inverse matrix $\mathbf{V}^{-1}$.   To avoid them, Tikhonov's regularization [3] was performed by replacing **V** with its regularized matrix $\mathbf{V}_{reg}$ as follows:

$$\mathbf{V}_{reg} = \mathbf{V} + \lambda \mathbf{I} \tag{13}$$

We empirically selected $\lambda = 1.0$ for this study.

Note that the calculation of Eq. (5) can be speeded up by using the following formula (derived from the Woodbury matrix identity),

$$\mathbf{V}_{reg}^{-1} = (\mathbf{V} + \lambda \mathbf{I})^{-1} = (\mathbf{U}^t \mathbf{D} \mathbf{U} + \lambda \mathbf{I})^{-1} = \lambda^{-1} \mathbf{I} - \lambda^{-2} \cdot \mathbf{U}^t (\mathbf{D}^{-1} + \lambda^{-1} \cdot \mathbf{I}')^{-1} \mathbf{U} \tag{14}$$

Here, **D** is a $(N\text{-}1) \times (N\text{-}1)$ diagonal matrix whose diagonal elements are nonzero eigenvalues of **V**, and **U** is a $(N\text{-}1) \times {}_MC_2$ matrix whose row vectors are corresponding eigenvectors.   **I** and **I'** are identity matrices whose sizes are $_MC_2$ and $N\text{-}1$, respectively.

2.2.2. Combinatorial optimization.   The maximization of Eq. (12) for all combinations of $\mathbf{X} = \{\mathbf{x}_m\}$, $\mathbf{x}_m \in \{\mathfrak{I}, \mathbf{c}_m^1, \ldots, \mathbf{c}_m^{N_m}\}$ is performed by a Gibbs' sampler-based simulated annealing algorithm reported by Geman & Geman [4].   In the algorithm, a virtual temperature T is introduced to modify the probability distribution. The distribution in (12) was modified as follows:

$$p_{\text{modified}}(\mathbf{X}; T) = \frac{1}{Z} \cdot \left[\left\{ \prod_{m=1}^{M} p(\mathbf{x}_m \mid \mathbf{I}_m) \right\} \cdot p(\mathbf{X}) \right]^{-\frac{1}{T}} \tag{15}$$

Z is a normalization factor in order to make the sum of probability 1.

When $T=1$, the modified distribution equals to the original in (12).   With higher T, the modified distribution is almost homogenous for any **X** in its domain.   However, it becomes sharper and more local with lower T.   In the limiting case of $T \to +0$, the modified distribution has nonzero probability only at its global maximum point. (Fig. 2)
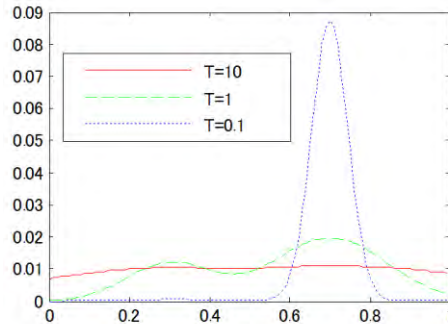
Fig. 2. A simple example how the probability distribution is modified by the temperature. The green curve illustrates an original distribution (a unary mixture Gaussian distribution for example). The modified distributions with $T = 10$ and 0.1 are illustrated with red and blue curves, respectively. Note that the distribution becomes flat with higher temperature, while it becomes sharper with lower temperature.

In the simulated annealing, each $\mathbf{x}_m$ ($m = 1, 2, …, M$) is sequentially and repeatedly sampled from the modified distribution by a Gibbs' sampler. The sampling begins with very high T, which decreases gradually, and finally it gets so low that the system converges to the maximum point. In this study the simulated annealing were performed for 100 cycles with T=1000 (in order to cancel the effect of the initial condition; so-called *burn-in*), then it was gradually cooled down to T=0.01 through 1000 cycles.

2.2.3. Handling of vertebral anomalies. A majority of human beings have 12 thoracic and 5 lumbar vertebrae. Segmentation anomaly of the spinal column is a common anatomical anomaly in which the subject has 11 or 13 thoracic, and/or 4 or 6 lumbar, vertebrae. The prevalence is, in a report, about 9 % [5]. This anomaly is very problematic in both defining and detecting vertebral landmarks (fig. 1a). The "state $\Im$" approach described in chapter 2.1. is not enough for them, because it is not only a local banishment of a single anatomical entity but causing a global morphological change in the spine.

To overcome this, a series of "anomaly landmark position set converters" are introduced. One converter can convert any landmark position set in a subject with a certain type of anomaly (e.g., 6 lumbar vertebrae, or "6L") into a virtually normalized landmark position set (i.e., a landmark position set as if she or he has only 5 lumbar vertebrae). It is simply performed by replacing each landmark coordinates by an appropriate internally dividing point between two of them (fig. 3). Because 7 types of anomalies (11T, 13T, 4L, 6L, 11T+6L, 13T+4L) were considered in this study, 7 different converters were designed. Through one of these converters, abnormal spines can be converted into a "normalized" one, with which the prior probability $p(\mathbf{X})$ can be calculated.
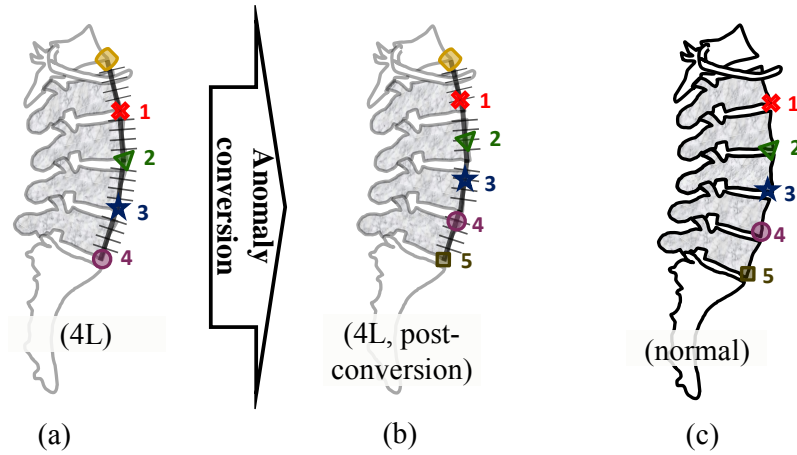
Fig. 3.   The anomaly conversion.   In this example, (a) a landmark position set in a case of 4-lumbar vertebra (4L) anomaly is virtually converted into (b) a "normal" landmark set with 5 vertebrae.   With this conversion, the posterior probability $p(\mathbf{X})$ of the given landmark position set can be calculated in the same manner as in the cases with (c) normal 5-lumbar spines.

Because the algorithm do not know which anomaly is correct, it must estimate it. It can be done by (i) hypothesizing one anomaly (or normal), (ii) calculating the posterior probability under the hypothesis and (iii) comparing the probabilities between all hypotheses.   In detail, two different strategies were evaluated:

(*i*) *Comparison after all optimization*.   Firstly, a series of combinatorial optimization with all converters are performed.   Then, the converter with the largest probability is chosen as the estimated anomaly (or normal).

(*ii*) *Comparison in situ*.   The comparison is performed at the end of every cycle in the simulated annealing.   In other words, the state "which converter is currently selected" is also dealt with as one extra variable to be optimized.

2.3. Evaluation

2.3.1. Virtual detector output construction.   The detector outputs were virtually created for each landmark and for each CT dataset.   If the target landmark existed in the subject's real body, it was included as a candidate, as well as 25, 50, 75 or 100 of false positive candidates.   Each false positive candidate was determined randomly following a 3-D Gaussian probability distribution (the center corresponds to the true point, and the standard deviations were $\sigma_x = \sigma_y = \sigma_z = \frac{1}{\sqrt{3}} \cdot 100$ millimeters).   Any false candidates within 20 millimeters from the true point were removed.   If the target landmark did not exist in reality, only false positive points were added, using another adjacent landmark as the distribution center.   The detector-estimated probability $p_m{}^{\mho}$ was fixed to be 0.05, and all of $p_m{}^i$ were set to be uniform.

2.3.2. Experimental settings.   The experiments were performed on a workstation with an Intel® Core™ i7-2600 processor and one NVIDIA® Tesla™ C2050 GPU

computing processor.   The latter was utilized to speed up the calculation.   The processing times were 9.7 minutes per case for *comparison after all optimization* strategy and 3.6 minutes for *comparison in situ* strategy with 100 candidates per landmark.

5 times of experiments were performed for each experimental setting, using different sets of virtual detector outputs.

The statistical model of squared distances (the mean $E(d_{i,j})$, variance $V(d_{i,j})$ and the covariance matrix **V**) was calculated by leave-one-case out method.

3   Results and discussion

The proposed algorithm was evaluated with 156 bony landmark points which were manually inputted in 50 thin-slice clinical human body CT datasets.   The list of landmark used is available in [2].   Among the 50 datasets, total 9 had a segmentation anomaly (three 13T, one 11T, three 6L, two 4L and none of 13T+4L or 11T+6L).

The summary of results is shown in Table 1.   An example result is also shown in Figure 4.   The accuracies of anomaly estimation were from 87.6 to 97.6%, varying among strategies and number of candidates.   The accuracy of anomaly estimation was better in *comparison after optimization* strategy than in *comparison in situ* strategy.   Though the former strategy takes approximately 3 times longer time, it seems to be useful especially for a larger number of candidates.

Table 1.   The entire result of anomaly estimation and combinatorial optimization.

	No. of cands	anomaly estimation accuracy (%)			landmark determination: existent-in-reality LMs (%)			non-existent LMs (%)	
		anomaly cases	normal cases	**overall**	TP	FN	FP$_{candidate}$	TN	FP$_{existence}$
comparison after opt.	25	88.9 ±13.6	99.5 ±1.1	**97.6 ±3.3**	96.60 ±2.90	0.09 ±0.13	3.31 ±2.78	100 ±0	0 ±0
	50	88.9 ±7.9	99.5 ±1.1	**97.6 ±1.7**	96.25 ±0.34	0.08 ±0.05	3.67 ±0.32		
	75	86.7 ±14.5	99.0 ±1.3	**96.8 ±2.7**	94.40 ±0.39	0.09 ±0.08	5.51 ±0.36		
	100	77.8 ±11.1	94.6 ±4.0	**91.6 ±3.0**	91.33 ±1.03	0.26 ±0.09	8.42 ±1.00		
comparison in situ	25	84.4 ±12.7	100 ±0	**97.2 ±2.3**	97.65 ±0.36	0.09 ±0.08	2.26 ±0.32		
	50	73.3 ±18.6	100 ±0	**95.2 ±3.4**	95.67 ±0.70	0.17 ±0.12	4.15 ±0.59		
	75	64.4 ±19.9	99.5 ±1.1	**93.2 ±3.9**	93.37 ±1.02	0.24 ±0.15	6.40 ±0.91		
	100	48.9 ±14.9	96.1 ±2.8	**87.6 ±2.6**	89.64 ±0.61	0.40 ±0.05	9.96 ±0.57		
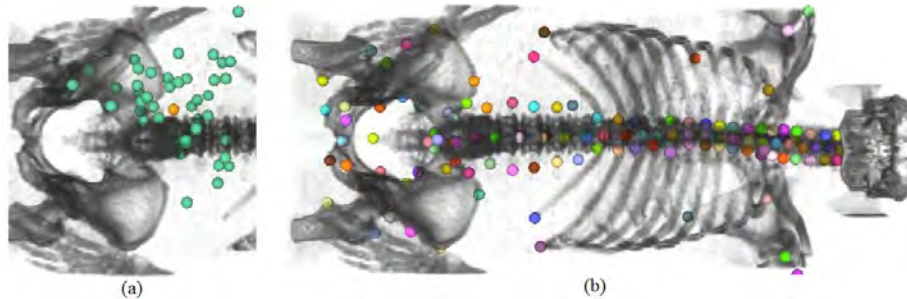
Fig. 4. An example result of a case with 6 lumbar vertebra (6L) anomaly.   (a) The false (green) and true (orange) candidates outputted by the virtual detector for the tip of right transverse process of the 4th lumbar vertebra.   (b) The optimization result for 156 landmarks.

The optimization result of each landmark in each case was also evaluated.   Any landmark which existed in reality was classified as one of TP, $FP_{candidate}$ or FN. Within them, TP means that the algorithm correctly selected the true position candidate. $FP_{candidate}$ means that the algorithm selected any false position candidate. If the algorithm concluded that the target landmark did not exist, but in reality it existed, it was categorized as FN.   In this study, the accuracy ratios (ratios of TP) for existent landmarks were varied from 89.64 to 97.65%.

Additionally, all of non-existing landmarks were classified as either TN or $FP_{existence}$.   TN means that the target landmark did not exist in reality and the algorithm correctly found it (either by adopting some anomaly hypothesis which does not include the target landmark, or by selecting the state $\Im$ for the landmark.)   If the algorithm chose any candidate other than $\Im$, it was classified as $FP_{existence}$.   In this study all of non-existent landmarks were determined as TN, so the accuracy was 100%.

It is noticeable that, throughout all experimental condition, none of non-existing landmarks were mistaken as $FP_{existence}$.   In fact, without any exception, all mistakes in anomaly estimation were underestimation of thoracic / lumbar vertebral number.   It may imply that our algorithm was not sufficiently optimized for the problem, favoring anomalies having less vertebral number, and yet to be investigated in the future work. Also, we are now planning to overcome this limitation by adding some additional term, which evaluates the regularity of vertebral bones' alignment, to the MAP estimation.

4   Conclusion

A novel combinatorial optimization algorithm for landmark detection and anomaly estimation was presented.   The proposed method showed fair results even with a large number of landmark position candidates.   Additionally, its feasibility to estimate segmentation anomaly of the vertebrae, which is one of the most common and problematic anomalies in detecting bone landmarks, has been shown.   Therefore, we believe that our algorithm is useful in medical image analysis such as a pre-

process for computer-assisted detection/diagnosis (CAD) applications. The future work will include improvement of accuracy in anomaly estimation and evaluation with real detector outputs with a large number of datasets.

Acknowledgement

References

1. Seifert, S., Barbu, A., Zhou, S.K., et. al.: Hierarchical parsing and semantic navigation of full body CT data. In: Pluim, J.P.W., Dawant, B.M. (eds.) SPIE (2009)
2. Nemoto, M., Masutani, Y., Hanaoka, S. et. al.: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. In: Proc. SPIE 2011, 7962--121 (2011)
3. Crimi, A., Sporring, J., Bruijne, M., et. al.: Prior knowledge regularization in statistical medical image tasks. Probabilistic Models for Medical Image Analysis 2009, MICCAI, Imperial College London, UK (2009)
4. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.6, no.6, 721--741 (1984)
5. Carrino, J. A., Campbell, P. D., Lin, D.C., et al.: Effect of spinal segment variants on numbering vertebral levels at lumbar MR imaging. Radiology, 259:203--212 (2011)