Landmark Localization in CT Images Using Dense Matching of Graphical Models

Vaclav Potesil[1]

Timor Kadir[1]

Günther Platsch[2]

Mike Brady[1]

[1] University of Oxford
Dept. of Engineering Science
Oxford, UK

[2] Siemens plc (Healthcare)
Molecular Imaging
Oxford, UK

Abstract

We present a method based on graphical models for the localization of corresponding anatomical landmarks in CT images of multiple patients for which only limited labeled training data is available. Our method mobilizes anatomical spatial relationships learnt from labeled training images in order to improve dense matching using weak landmark appearance descriptors. In this study, we report results for localization of 22 different anatomical landmarks in 20 unseen lung cancer patients and different types of anatomical constraints (none, box-range, Gaussian). The average registration error over all landmarks improved from 18.8 voxels (37.6mm) of the raw landmark descriptors to 4.2 voxels (8.4 mm) using the anatomical constraints.

1   Introduction

Inter-subject matching and registration of whole-body oncology CT images is a challenging problem due to the intrinsically high variability of normal subjects and of pathological structures. The motivation for our work is to develop algorithms for improved inter-subject registration in whole-body PET/CT oncology applications. In this work, we report methods to match corresponding structures of multiple patients for which there is only a limited labeled training dataset. Our approach is based on a parts-based graphical model.

Conventional approaches to determining correspondences in medical imaging typically rely on registration methods. However, despite considerable advances in deformable registration, there still do not exist reliable methods for aligning whole-body images of different subjects. The performance of inter-subject registration could be improved by informative priors that capture the wide variability of structures. Active shape/appearance models (ASM/AAM) have addressed this problem [2]. However, several authors have drawn attention to the limitations of such global models when applied to clinical images in which there are significant local abnormalities [1, 7]. Graphical Models , developed in computer vision, offer an alternative approach to modeling flexible objects, one which does not impose explicit global priors, and which is generally considered to be potentially more robust to local abnormalities than global AAMs.

In medical image analysis, the attractive properties of graphical models have attracted attention in spine labeling [3, 6]. However, whole-body matching involves considerably more complex anatomical variability. Here, we report a method that is capable of localizing a broad range of structures of interest to a clinician, including both skeletal and soft tissues

2 Methods

Our method is based on the Pictorial Structure model [4, 5], which we summarize here.

2.1 Pictorial Structure Representation

All landmarks share the same representation, which comprises a unary energy term for part appearance and a set of pair-wise terms for spatial compatibility. It takes the form of a tree-structured undirected graphical model G with nodes v_i, v_j representing the landmarks.

$$P(L \mid I, \Theta) \propto \left(\prod_{i=1}^{n} p(I \mid l_i, u_i)^{\alpha} \prod_{v_i, v_j \in G} p(l_i, l_j \mid c_{ij}) \right) \tag{1}$$

In this equation, the first product term represents the cost of each landmark v_i at location l_i, given the quality of match of landmark appearance model u_i to the image I. The second term is the compatibility cost of connected pairs of landmarks assuming conditional independence (i.e., anatomical constraints, represented by c_{ij}). This equation can be rewritten as an energy minimization problem by taking its negative logarithm. L is the vector of locations l_i assigned to all landmarks v_i, Θ are the model parameters learnt from the data. α is a normalization constant, selected empirically for each type of spatial constraint and fixed for all parts (for box-range constraints: $\alpha = 1$, for Gaussian constraints: $\alpha = 20$).

Local Appearance. We model local tissue appearance in terms of local rectangular image patches around a candidate landmarks with a fixed scale h=12 voxels. Patches are projected onto a set of n = 17 bases that includes the mean template and the top n = 16 principal eigen-patches obtained from ground truth patches for each landmark. Additional positive examples were generated from the hand-annotated patches by random rotations within anatomically plausible ranges. Negative examples of non-landmark tissue ("background") are sampled from unlabelled tissue from a spherical region of interest within the segmented body outline. The variability of part appearance is represented by the distribution of the coefficients, $u_i \sim N(M, \Sigma)$ for each part i, where M, Σ are diagonal matrices. $u_{bi} \sim N(M, \Sigma)$ is the model for the local "background".

Pair-wise Anatomical Constraints. Pair-wise spatial compatibility terms $G_{ij}(x_i, x_j)$ penalise part placements outside of the degree of anatomical variability exhibited within the training database The functional form of this spatial model is approximated as a distribution over $\vec{l}_{ij,k} = l_{i,k} - l_{j,k}$ (the relative position vector of two parts i, j in training patient k). We evaluate: no constraints; box-range constraints $P(l_i \mid l_j) \sim U\left(\min \vec{l}_{ij,k}, \max \vec{l}_{ij,k} \right)$; (search within a bounding cube); and Gaussian models $P(l_i \mid l_j) \sim N\left(\vec{l}_{ij,k}, \mu_i, \Sigma_i \right)$ with diagonal and full covariance. The Gaussian terms are truncated at 3 standard deviations.

2.2 Learning Pictorial Structure Parameters

The parameters of the model are learnt as maximum likelihood (ML) estimates from the training data D. The connectivity of graph G is obtained based on an estimated n-by-n matrix Q of edge compatibilities, $Q_{ij}(G \mid D) \sim \sum_{l_i, l_j \in D} -\log P(l_i, l_j \mid c_{ij})$.

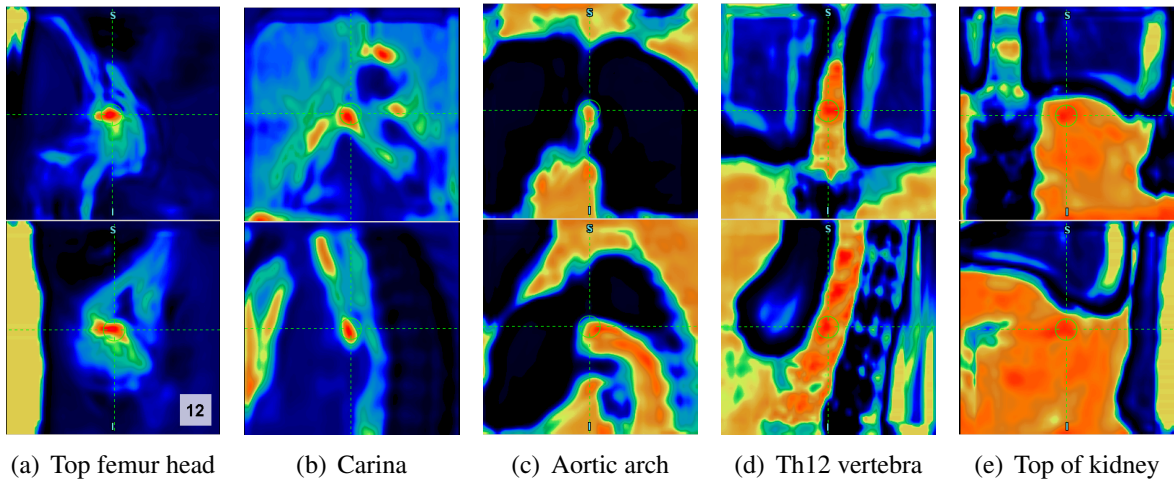| (a) Top femur head | (b) Carina | (c) Aortic arch | (d) Th12 vertebra | (e) Top of kidney |

Figure 1: Example landmark posterior probability maps. Top: Coronal, Bottom: Sagittal views. Images show a 16x16x16 cm region centered at the ground-truth location. Bright colors correspond to high probability landmark placements (1). Dark colors correspond to low probability landmark placements (0). White box indicates patch scale h = 12 vox.

A set of edges connecting all parts in a tree is obtained by finding the minimum spanning tree of the complete graph such that the sum of edge costs is minimum. Intuitively, this can be seen as searching for edges with the most compact ("rigid") pair-wise constraints.

The pair-wise spatial model is fitted by directly calculating the MLE estimate for the vector $\vec{l}_{ij,k}$ (i.e. the mean and the diagonal or full covariance matrix for the Gaussian or bounding cube for the box-range model). The appearance model parameters are obtained by direct calculation of the sample mean and variance of the feature coefficients.

2.3 Matching Pictorial Structure to New Image

To find the best placements for the parts in a new image, we apply the appearance models to the whole image using a sliding window approach. The log-likelihood ratios are combined using a Naive Bayes classifier to obtain the posterior probability for each landmark and sliding window placements. For computational reasons, the dense descriptors were evaluated in a cube-shaped region of interest centered around the ground-truth landmark placement ([81 x 81 x 81 voxels]).

Min-sum belief propagation algorithm is used for an exact, globally optimal fitting using the (dense) appearance posterior probability maps and the pair-wise compatibility terms. Pair-wise terms are evaluated using the fast generalized Distance transforms [4] and fast min/max filters.

3 Data and Results

To date, our database comprises 83 lung cancer PET/CT cases (here, randomly split into 63 training and 20 unseen test images). Contrast-enhanced, diagnostic-quality CT scans of the torso were acquired using Siemens Biograph 6 and re-sampled to 2mm isotropic resolution.

22 clinical landmarks were selected by an expert radiologist with 20 years experience, according to their utility as anatomical reference points for whole-body PET/CT. Ground-truth landmark positions were annotated by a non-expert reader, who followed interactive guidance by the expert. The expert subsequently validated a subset of the annotations, including a) All placements flagged as uncertain or abnormal b) All annotations where the non-expert disagreed with another non-expert c) All 30% of randomly selected patients.
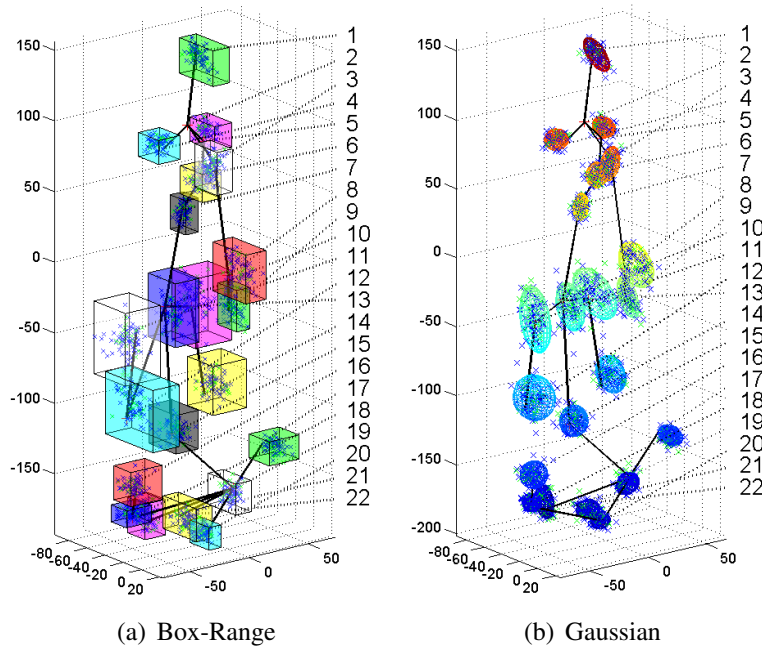
(a) Box-Range (b) Gaussian

Figure 2: Anatomical constraints. Colored shapes show valid constraint ranges, overlaid on one training patient (black lines). Blue and green points represent the training and test points respectively, plotted in the local coordinates of the neighboring landmarks. Labels correspond to landmarks 1 - 22 (Table 1)

Performance was assessed in terms of the robustness to misclassification and landmark registration accuracy with different spatial constraint models. Localization accuracy was measured in terms of the mean RMS error. Robustness was measured as the proportion of failed detections, defined as parts placed above 10 voxels from the ground truth landmark location (Table 1).

4   Discussion and Conclusion

We aim to establish correspondence for a limited set of clinical anatomical landmarks which are guaranteed to be present and localizable in most clinical oncology images. Our method can compensate for weak appearance descriptors with a high level of false positives and can disambiguate repeated self similar structures by mobilizing relations between landmarks in the image.

Our experiments show that all types of anatomical constraints improve registration compared to using the appearance descriptors on their own. Localization of most landmarks improved significantly, as evidenced by the reduction of average Mean RMS error from 18.8 voxels down to 4.2 voxels and the reduction in misclassifications from 48% down to 8%. Gaussian constraints consistently outperform box-range constraints. However, we saw no improvement on average from using full over diagonal covariance, which suggests the Gaussian distributions may be a wrong model for some relations.

Some landmarks remain relative poorly localized. These include bottom sternum, a highly variable structure with ambiguous ground-truth placements and the center of bladder, an intrinsically poorly localizable structure, due to variable fullness and appearance of any contrast agent. Moreover, in coccyx and top of kidney, high variability of both appearance and spatial relations hinders more accurate localization. Finally, Th12 vertebra cannot be accurately localized with the current model as the available constraints are too broad to

Table 1: **A. Landmark registration accuracy.** The table shows mean RMS error from the ground-truth location. Lower error means more accurate localization. **B. Landmark detection robustness.** The table shows failure rate (% detections above 10 voxel RMS). Lower failure rate means more reliable localization. Results shown for 22 landmarks and four constraint types (none, box-range, diagonal and full Gauss).

		A) Avg. RMS Error [Voxels]				B) Misclassification rate [% Failed detections]			
		No	Box	Diag	Full	No	Box	Diag	Full
1	C2 vertebra	27.4	8.0	**1.9**	**1.9**	75	20	**0**	**0**
2	C7 vertebra	11.3	4.9	2.5	**2.3**	40	5	**0**	**0**
3	top of the sternum	16.3	5.2	3.1	**2.8**	45	15	**0**	5
4	top right lung	15.7	3.2	**3.6**	3.7	40	**0**	**0**	**0**
5	top left lung	3.1	3.1	3.0	**2.9**	**0**	**0**	**0**	**0**
6	aortic arch	12.9	4.9	4.1	**4.0**	35	5	**0**	**0**
7	carina	22.7	2.3	2.4	**2.3**	40	**0**	**0**	**0**
8	low sternum (ribs)	28.8	11.4	6.6	**6.2**	95	50	15	**10**
9	low sternum (tip)	26.3	12.6	9.1	**8.4**	80	50	35	**30**
10	Th12 vertebra	17.9	7.5	**5.7**	6.0	80	40	**25**	25
11	top right kidney	37.0	10.2	**4.6**	4.9	75	35	**15**	20
12	bottom right kidney	6.7	2.6	**2.8**	2.9	10	**0**	**0**	**0**
13	top left kidney	32.8	7.2	**6.5**	**6.5**	80	25	**20**	**20**
14	bottom left kidney	15.0	2.8	**3.6**	3.7	30	**5**	10	**5**
15	L5 vertebra	19.6	6.0	**5.0**	5.1	70	25	**20**	20
16	right illiac crest	15.7	4.9	3.9	**3.8**	30	10	5	**0**
17	left illiac crest	10.8	3.3	4.0	**3.9**	20	**0**	**0**	**0**
18	right head of femur	9.8	2.1	2.2	**2.2**	20	**0**	**0**	**0**
19	left head of femur	4.4	2.8	**2.4**	2.6	5	**0**	**0**	**0**
20	symphysis	19.6	3.9	**3.3**	3.4	50	**0**	**0**	**0**
21	os coccygeum	15.1	4.9	**4.1**	4.9	35	15	**10**	15
22	center of bladder	43.7	17.0	**8.6**	**8.6**	95	80	40	**30**
	Average	18.8	5.9	**4.2**	**4.2**	48	17	9	**8**

disambiguate from unlabeled L1 and Th11 vertebra. Overall, the Gaussian model appears better suited than box-range. However, it introduces a certain bias to "population mean" and box constraints may be more suited if a better appearance descriptors are available. We used an ad-hoc generative appearance descriptor and didn't attempt to optimize the parameters (anisotropic patch scale, number of features etc.), which we leave for future work. The performance achieved to date in our experiments is also limited by the spatial resolution of the down-sampled data as well as the variability in the ground-truth annotations, particularly for the intrinsically less well localized landmarks such as the top of the lung. Moreover, here we use only one tree-structured model to cover all the landmarks. Locally optimized graph structure may provide improved spatial constraining to a particular landmark of interest.

Acknowledgements

References

[1] R. Beichel, H. Bischof, F. Leberl, and M. Sonka. Robust active appearance models and their application to medical image analysis. *IEEE Transactions on Medical Imaging*, 24 (9):1151–1169, 2005.

[2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[3] J. J. Corso, RS Alomari, and V. Chaudhary. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In *MICCAI*, volume 11, page 202, 2008.

[4] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.

[5] MA Fischler and RA Elschlager. The representation and matching of pictorial structures. *Computers, IEEE Transactions on*, 100(22):67–92, 1973.

[6] M. Bergtholdt et al. S. Schmidt, J. Kappes. Spine detection and labeling using a parts-based graphical model. *LECTURE NOTES IN COMPUTER SCIENCE*, 4584:122, 2007.

[7] M. Toews and T. Arbel. A statistical parts-based model of anatomical variability. *Medical Imaging, IEEE Transactions on*, 26(4):497–508, 2007.