Accurate Intervertebral Disc Localisation and Segmentation in MRI Using Vantage Point Hough Forests and Multi-atlas Fusion

Mattias P. Heinrich[1(✉)] and Ozan Oktay[2]

[1] Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
heinrich@imi.uni-luebeck.de
[2] Biomedical Image Analysis Group, Imperial College London, London, UK
http://www.mpheinrich.de

Abstract. An accurate method for localising and segmenting intervertebral discs in magnetic resonance (MR) spine imaging is presented. Atlas-based labelling of discs in MRI is challenging due to the small field of view and repetitive structures, which may cause the image registration to converge to a local minimum. To tackle this initialisation problem, our approach uses **Vantage Point Hough Forests** to automatically and robustly regress landmark positions, which are used to initialise a discrete deformable registration of all training images. An image-adaptive fusion of propagated segmentation labels is obtained by non-negative least-squares regression. Despite its simplicity and without using specific domain knowledge, our approach achieves sub-voxel localisation accuracy of 0.61 mm, Dice segmentation overlaps of nearly 90% (for the training data) and takes less than ten minutes to process a new scan.

1 Introduction and Related Work

Automatic analysis of vertebras intervertabral discs in clinical 3D volumes of the spine is useful for diagnosis, monitoring of disease progression, image-guided surgical interventions and population studies [16]. While segmenting and localising vertebra bodies have been predominantly performed in CT scans [8,9,15], the soft tissue contrast and non-ionising acquisition of magnetic resonance imaging (MRI) makes it the preferred modality for intervertabral disc analysis [2,3]. Automated analysis of spine images, which has seen increased research interest over the last years (also due to the *SpineWeb*[1] initiative), is challenging due to the repetitive appearance of vertebras, restricted field-of-views. Therefore applying standard segmentation propagation approaches (multi-atlas segmentation) can easily fail [9] and/or require very long processing times. Thus model-based approaches [15], the integration of graphical model information [4,19] as well as regression forests [2,3,8] have been employed to increase robustness for finding and labelling the correct structures. The goal of the *"Automatic Intervertebral Disc Localization and Segmentation from 3D Multi-modality MR (M3) Images"*

[1] http://spineweb.digitalimaginggroup.ca.

J. Yao et al. (Eds.): CSI 2016, LNCS 10182, pp. 77–84, 2016.
DOI: 10.1007/978-3-319-55050-3_7

challenge held in conjunction with MICCAI 2016 is the identification, localisation and segmentation of seven discs, which are mainly within the lumbar spine.

Our work follows a similar methodological approach as [2,9]. First, we robustly localise the 3D position of the seven discs using a combination of a regression forest, Hough accumulation and a graphical model. Second, we use these positions to initialise a fast, discrete multi-atlas registration framework, which is followed by a non-negative least-squares regression of the most likely segmentation label. These 3D segmentations are then employed to refine the localisation estimation. While we build upon previous work, our approach contains elements that are (to the best of our knowledge) new to automatic spine segmentation/localisation. First, instead of using a supervised axis-parallel regression forest we adapt the recent concept of vantage-point forests [12] for regression, which has been shown to outperform random forests for multi-organ segmentation. Here, the whole length of a binary context feature vector is used to cluster the data meaningfully without being reliant on ground truth information during tree generation. Second, the combination of very fast deformable registration [10] and regression-based label fusion algorithms [14] enables processing times of less than ten minutes for a multi-atlas label fusion (MALF) reducing the time requirements compared to most state-of-the-art approaches substantially.

The paper is outlined as follows: we begin by describing our vantage point regression forests in Sect. 2, which is followed by a vote accumulation in Hough space and a simple spatial regularisation of candidates using a graphical (Markov chain) model for an accurate prediction of all disc centres in a new unseen scan. Note, that our approach does not make any specific use of domain knowledge and would therefore be applicable to other anatomical localisation tasks. Afterwards, the multi-atlas registration and label fusion framework is presented based on [10,14] in Sect. 3. A detailed flow-chart of all algorithmic steps is presented in Fig. 1. Finally, we present our experiments and results on the training dataset of the challenge in Sect. 4 and discuss our conclusion in Sect. 5. Note, that currently

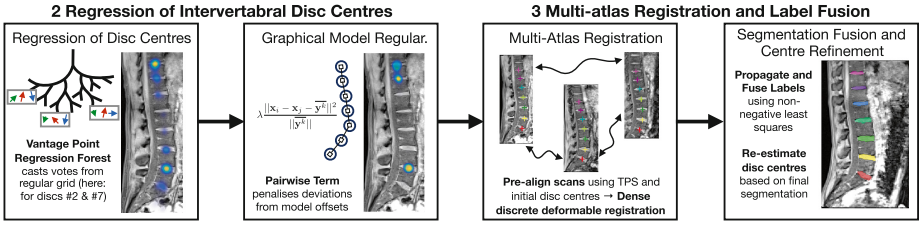


Fig. 1. Flow-chart of our proposed algorithm for accurate intervertabral disc localisation and segmentation. First, initial locations of disc centres are robustly found using vantage point forests and a graphical model. Next, the disc centres are used as known correspondences to initialise a deformable registration using a thin-plate spline (TPS) warp. Finally, multi-atlas label fusion is performed for accurate voxelwise segmentation and a refinement of disc centres.

we only employ the standard proton MRI sequence, but further improvements are to be expected when using all multi-modal scans.

2 Regression of Intervertabral Disc Centres

In order to analyse new scans completely automatically a robust initialisation of the correct disc positions is often necessary. A common problem for disc localisation in MR spine images is the confusion of two neighbouring discs due to their similarity and missing anatomical context (abdominal organs are not clearly visible in the field-of-view). We build our regression upon the recent concept of vantage point forests [12]. Since, a global localisation is sought, we sample patches on a uniform grid of locations $\mathbf{x}_i$ across the whole image domain. In training the position of seven ground-truth disc centres $\mathbf{y}_k$ is obtained as the centre of mass of the provided segmentation masks.

An intensity patch $\mathcal{P}_i \in \mathbb{R}^{|L|}$ (with L being the set of voxels), which is smoothed by a Gaussian kernel with variance σ_p^2, will be represented by a feature vector $\mathbf{h}_i \in \mathbb{H}^n$, where $h_{id} \in \{\pm 1\}$ defines the d-th dimension of the vector h_{id} corresponding to sample i. For this specific application, we restrict the feature values to be binary (in Hamming space $\mathbb{H}^n$) and can be simply obtained by a comparison of two random locations (q, r) within the patch:

$$h_{id} = +1 \text{ if } \mathcal{P}_i(q) > \mathcal{P}_i(r) \quad \text{for } (q,r) \in L \text{ and } h_{id} = -1 \text{ else} \tag{1}$$

as done in previous work on organ or keypoint localisation [1,17]. Note, that the same random sampling layout is used for every location. The use of binary features improves robustness against contrast variations often present in MRI scans [17]. The vantage point tree [20] is a data-structure that is suitable to cluster high-dimensional feature spaces into nested hyperspheres. In contrast to previous work on regression forests for landmark localisation [4,8], we **do not** perform supervised node optimisation but simply choose a random data point j from the current node (vantage point) for clustering as follows. The Hamming distance $d_H(i,j) = ||\mathbf{h}_i - \mathbf{h}_j||_{\mathbb{H}}$ (of the whole feature vectors) to all other data points i within the current node is calculated and the median distance τ is used as threshold to split the data into two equal-sized sets that form the left and right predecessor (child) nodes (see [12] for more details and an implementation). When reaching the leaf node a displacement vector $\mathbf{d}_i^k = \mathbf{y}^k - \mathbf{x}_i$ is stored for every sample i and every landmark $k \in \{1, 2, \ldots, 7\}$. Using the full binary feature vectors enables very discriminative splits even without explicitly modelling the distribution of displacement vectors and is computationally very efficient due to the implementation of the Hamming weight as `popcount` instruction in current CPUs [1]. A distance threshold $\delta_{\max}$ can be used to discard votes (during testing) from very far away locations. An ensemble of several randomly different vantage point trees is built to increase the generalisation.

During test the same random sampling layout is used as in Eq. 1 to extract binary feature vectors for a set of regular grid locations $\mathbf{x}_i$. After traversing each tree all training exemplars are collected and only the displacement vote of

the one ($\mathbf{d}_{i^*}$) with lowest Hamming distance with respect to the test sample is retained. Effectively, vantage point forests enable a very efficient approximate nearest neighbour search in Hamming space. The votes of all test locations (with offset vectors $\mathbf{x}_i + \mathbf{d}^k_{i^*}$) are accumulated in 7 Hough volumes $\mathbf{H}^k$ (one for each landmark, cf. [4,6]), which are later smoothed by a Parzen window kernel with σ_H. Finally, a graphical model [18] is used to impose spatial constraints and avoid confusions of neighbouring discs (which occurred twice for all 56 discs in training). Dynamic programming is applied to all possible pair-wise combinations of candidates of neighbouring discs. The unary term for the model is chosen to be the negative exponential of the accumulated Hough votes for any image location, which results in probability maps $\mathbf{H}^k$ of same size as the input image. The pairwise regularisation cost (weighted by λ) is the squared Euclidean distance between (average) model offset $\overline{\mathbf{y}^k} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i^k - \mathbf{y}_i^{k+1})$ and difference between the two respective locations:

$$E(\mathbf{x}_i, \mathbf{x}_j, k, \mathbf{H}) = \exp(-\mathbf{H}^k(\mathbf{x}_i)) + \lambda \frac{||\mathbf{x}_i - \mathbf{x}_j - \overline{\mathbf{y}^k}||^2}{||\overline{\mathbf{y}^k}||} \tag{2}$$

The minimum of $E(\mathbf{x}_i, \mathbf{x}_j, k, \mathbf{H})$ for each possible combination $(\mathbf{x}_i, \mathbf{x}_j)$ for two connected landmarks $(k, k+1)$ can be computed in linear complexity using distance transforms of sampled functions [5]. Marginal distributions of the likelihood (or vice-versa uncertainty) of the position of all landmark positions can be obtained following [11,19].

3 Multi-atlas Registration and Label Fusion

The publicly available non-parametric discrete registration tool **deeds** of [10] was used due to its computational efficiency and good results for MRI segmentation propagation. Given the estimated landmark localisations for a test scan (using the outcome of the previous section) and the ground truth information in training scans, we generate bounding boxes (using average disc sizes) and match a thin-plate spline transformation to their corner points. This transformation is used to pre-align all training images (and segmentation masks). Local cross correlation with a radius of $r = 3$ was used as similarity metric together with a Gaussian smoothing of 1.2 voxels and symmetry constraint for regularisation. The default multi-resolution and search range settings were used. Each deformable registration took around 60 s.

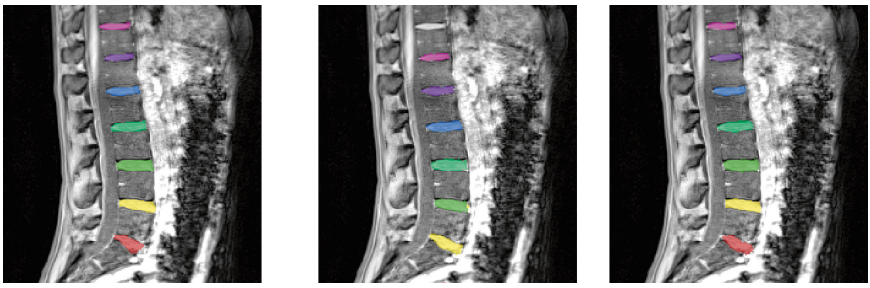
Following the well-known concept of multi-atlas label fusion (MALF), we estimate a local weighting for each (of the 7) registered atlas scan based on local cross correlation and a non-negative least square regression [14]. This step produces spatially coherent and accurate disc segmentations (see Fig. 2(c)), is very fast in practise (≈10 s), and can effectively compensate registration errors. Afterwards, the disc locations are re-estimated as the centre of mass of the fused segmentation labels for improved accuracy.

4 Experiments and Results on Training Data

In our experiments the combination of vantage point forests, Hough aggregation and this simple graphical model (see example results in Fig. 2) achieved very robust results (without a single misclassification of intervertebral discs) and required less tuning than random regression forests (for which we could not find sufficient settings for all training cases). We smoothed patches with $\sigma_p = 2.5$ mm, used $n = 320$ binary features drawn randomly within a radius of 25 mm, a stride of 4 voxels for the regular grid of voting voxels $\mathbf{x}_i$, and built 15 trees with a leaf size termination of 5. The application of the model to a test scan took approx. 2 s (including Hough aggregation and graphical model). The Parzen kernel for Hough aggregation was $\sigma_H = 3.75$ mm, the distance threshold $\delta_{\max} = 37.5$ mm and the regularisation weighting $\lambda = 2$. We obtained an average localisation error of 7.09 mm (max: 39 mm) without and 3.87 mm (max: 10 mm) with graphical model (see Table 1). While a lower error could easily be achieved by including a (cascaded) refinement stage [3,7], we are here mainly interested in the robustness of this step, since small misalignments will easily be corrected by the following deformable registration. After applying the multi-atlas registration and label fusion of Sect. 3, we achieve very high segmentation overlap (with an average Dice of 0.89) and a very low disc location error of 0.69 mm (the scan resolution is 1.25 mm^3) using the centre-of-mass re-estimation.

Table 1. Quantitative evaluation of our vantage point Hough forest regression (VPF). The robustness is increased by a subsequent graphical model (MRF). When used to initialise a fast multi-atlas registration and label fusion (MALF), very low localisation and segmentation surface distances as well as high Dice scores are achieved.

Method	Metric	avg.	#1	#2	#3	#4	#5	#6	#7	#8
MALF w/o regr.	Localisation (mm)	**5.51**	0.75	37.58	2.79	0.48	0.71	0.53	0.79	0.50
	Surface dist. (mm)	**3.34**	0.52	19.92	3.65	0.56	0.44	0.39	0.40	0.85
	Dice overlap	**0.76**	0.89	0.02	0.70	0.90	0.88	0.89	0.90	0.88
VPF+Hough	Localisation (mm)	**7.09**	4.10	12.44	12.60	9.25	3.44	2.52	5.58	6.80
	Loc. max (mm)		5.72	39.89	38.41	39.81	5.32	5.79	26.82	27.31
+MRF	Localisation (mm)	**3.87**	3.87	2.90	3.78	5.25	3.08	4.16	3.91	3.97
	Loc. max (mm)		5.51	7.07	6.38	8.93	5.50	7.86	7.68	10.05
+MALF	Localisation (mm)	**0.61**	0.76	0.57	0.52	0.53	0.68	0.50	0.84	0.50
	Surface dist. (mm)	**0.38**	0.37	0.39	0.39	0.38	0.36	0.36	0.35	0.39
	Dice overlap	**0.89**	0.89	0.89	0.90	0.91	0.89	0.89	0.90	0.88



(a) Overlay with ground truth (b) MALF with linear init. (c) MALF with VPF init.

Fig. 2. Sagittal slice of 3D MRI of *case B* with segmentations of all seven discs overlaid in colour. (a) Ground truth manual segmentation. (b) Misaligned segmentations using standard multi-atlas label fusion (MALF) due to poor initialisation. (c) Proposed vantage point forest regression improves overlap and successfully segments all discs as it provides better initialisation for MALF. Best viewed in colour. (Color figure online)

5 Discussion

We have presented a simple yet very robust and fast method for finding anatomical landmarks (intervertebral discs) in spine MRI scans. The use of (unsupervised) vantage point forest together with discriminative binary feature vectors enables very good regression results without tuning of different trade-offs between classification and regression in supervised random forests. A subsequent multi-atlas registration and label fusion (initialised using a thin-plate spline transform obtained from this automatic disc localisation) achieve a Dice score of 89% on average and a refined average localisation error of 0.69 mm with a processing time of ≈10 min per unseen scan. Further improvements may be obtained by employing all multi-channel MR sequences (here we only used the proton MRI), which could be easily integrated using [13]. The source-code for all processing steps will be made available on http://mpheinrich.de/software.html.

References

1. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: computing a local binary descriptor very fast. IEEE PAMI **34**(7), 1281–1298 (2012)
2. Chen, C., Belavy, D., Yu, W., Chu, C., Armbrecht, G., Bansmann, M., Felsenberg, D., Zheng, G.: Localization and segmentation of 3D intervertebral discs in MR images by data driven estimation. IEEE Trans. Med. Imag. **34**(8), 1719–1729 (2015)
3. Chen, C., Belavy, D., Zheng, G.: 3D intervertebral disc localization and segmentation from MR images by data-driven regression and classification. In: Wu, G., Zhang, D., Zhou, L. (eds.) MLMI 2014. LNCS, vol. 8679, pp. 50–58. Springer, Cham (2014). doi:10.1007/978-3-319-10581-9_7
4. Donner, R., Menze, B.H., Bischof, H., Langs, G.: Global localization of 3D anatomical structures by pre-filtered hough forests and discrete optimization. Med. Image Anal. **17**(8), 1304–1314 (2013)

5. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions. Theory Comput. **8**, 415–428 (2012)
6. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(11), 2188–2202 (2011)
7. Gauriau, R., Cuingnet, R., Lesage, D., Bloch, I.: Multi-organ localization with cascaded global-to-local regression and shape prior. Med. Image Anal. **23**(1), 70–83 (2015)
8. Glocker, B., Feulner, J., Criminisi, A., Haynor, D.R., Konukoglu, E.: Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 590–598. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_73
9. Glocker, B., Zikic, D., Haynor, D.R.: Robust registration of longitudinal spine CT. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 251–258. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_32
10. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Non-parametric discrete registration with convex optimisation. In: Ourselin, S., Modat, M. (eds.) WBIR 2014. LNCS, vol. 8545, pp. 51–61. Springer, Cham (2014). doi:10.1007/978-3-319-08554-8_6
11. Heinrich, M.P., Simpson, I.J., Papież, B.W., Brady, M., Schnabel, J.A.: Deformable image registration by combining uncertainty estimates from supervoxel belief propagation. Med. Image Anal. **27**, 57–71 (2016)
12. Heinrich, M.P., Blendowski, M.: Multi-organ segmentation using vantage point forests and binary context features. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 598–606. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_69
13. Heinrich, M.P., Papież, B.W., Schnabel, J.A., Handels, H.: Multispectral image registration based on local canonical correlation analysis. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 202–209. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_26
14. Heinrich, M.P., Wilms, M., Handels, H.: Multi-atlas segmentation using patch-based joint label fusion with non-negative least squares regression. In: Wu, G., Coupé, P., Zhan, Y., Munsell, B., Rueckert, D. (eds.) Patch-MI 2015. LNCS, vol. 9467, pp. 146–153. Springer, Cham (2015). doi:10.1007/978-3-319-28194-0_18
15. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in CT images. Med. Image Anal. **13**(3), 471–482 (2009)
16. Li, S., Yao, J., Navab, N.: Guest editorial special issue on spine imaging, image-based modeling, and image guided intervention. IEEE Trans. Med. Imag. **34**(8), 1625–1626 (2015)
17. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast multiple organ detection and localization in whole-body MR dixon sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23626-6_30
18. Potesil, V., Kadir, T., Platsch, G., Brady, M.: Improved anatomical landmark localization in medical images using dense matching of graphical models. In: BMVC, vol. 4, p. 9 (2010)

19. Richmond, D., Kainmueller, D., Glocker, B., Rother, C., Myers, G.: Uncertainty-driven forest predictors for vertebra localization and segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 653–660. Springer, Cham (2015). doi:10.1007/978-3-319-24553-9_80
20. Yianilos, P.N.: Data structures and algorithms for nearest neighbor search in general metric spaces. In: SODA 1993, pp. 311–321 (1993)