Multi-Organ Segmentation using Vantage Point Forests and Binary Context Features

Mattias P. Heinrich and Maximilian Blendowski

Institute of Medical Informatics, University of Lübeck, Germany
heinrich@imi.uni-luebeck.de, www.mpheinrich.de

Abstract. Dense segmentation of large medical image volumes using a labelled training dataset requires strong classifiers. Ensembles of random decision trees have been shown to achieve good segmentation accuracies with very fast computation times. However, smaller anatomical structures such as muscles or organs with high shape variability present a challenge to them, especially when relying on axis-parallel split functions, which make finding joint relations among features difficult. Recent work has shown that structural and contextual information can be well captured using a large number of simple pairwise intensity comparisons stored in binary vectors. In this work, we propose to overcome current limitations of random forest classifiers by devising new decision trees, which use the entire feature vector at each split node and may thus be able to find representative patterns in high-dimensional feature spaces. Our approach called vantage point forests is related to cluster trees that have been successfully applied to space partitioning. It can be further improved by discarding training samples with a large Hamming distance compared to the test sample. Our method achieves state-of-the-art segmentation accuracy of ≥90% Dice for liver and kidneys in abdominal CT, with significant improvements over random forest, in under a minute.

Keywords: Hamming space, random forests, patch-based classification

1 Introduction

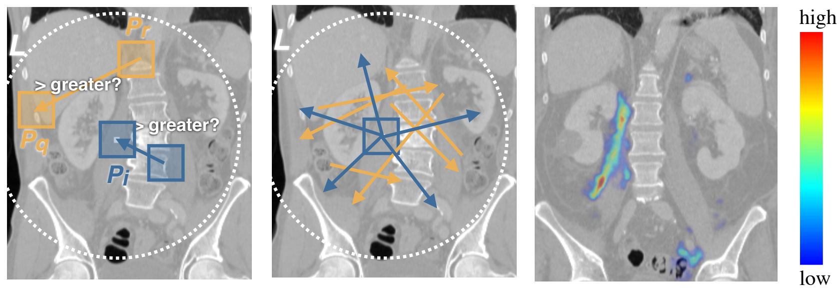
Multi-label classification is important in a number of vision applications, such as object recognition and medical image segmentation [1]. Clinical applications of automatic multi-organ segmentation may require either highly accurate delineations (e.g. for diagnostic tasks) or fast and robust segmentations of multiple organs (e.g. for image guided interventions). Manual interaction is often not possible in time-sensitive scenarios. We are hence interested in supervised segmentation, where a representative training set is available with dense manual label annotations, which is used to classify voxels in an unseen image based on features that are extracted within their spatial proximity (patches). Since time-efficient classifiers are required for large volumes, random forests [1, 2], fern ensembles [3, 4] or atlas forests [5] have become popular for localising anatomical landmarks or voxelwise multi-organ segmentation of medical scans. The quality of the segmentation might be limited due to changes in contrast or strong noise within the given scans and the availability of enough training samples (which may reduce the generalisability of certain classifiers). So far, the segmentation quality of multi-atlas based registration combined with label fusion (MALF) often outperforms voxelwise classification. These methods benefit from strong contextual correlations (in human anatomy) by using constrained transformation models, but the deformable image registration is usually very time-consuming.

The goal of this work is to improve segmentation quality (close to the level of registration-based approaches) while retaining the low computation times of ensemble classifiers. To deal with the above challenges we propose to use binary vectors of contextual features together with a newly developed tree-based classifier. To capture contextual information a 3D extension of BRIEF [6] (a popular 2D keypoint descriptor) is used, which is based on voxel comparisons within a (pre-smoothed) patch. In contrast to the related long-range context features introduced e.g. in [7] or SIFT vectors (used for medical segmentation in [8]), only the sign of intensity differences is stored, which makes the feature vectors robust against mononotic changes in intensities (in most medical scans orientational invariance is of lesser importance). Details of the exact sampling layout of the features used in this work will be given in Sec. 2.1. Our hypothesis is that the joint combination of hundreds of these BRIEF features can successfully capture contextual anatomical information and variability, but **only if** they are employed within a strong classifier.

Our approach for devising such a suitable classifier (that can deal with high-dimensional binary vectors) is based on the adaption of the vantage point trees, which was originally proposed for high-dimensional data clustering and accelerated search in metric spaces [9], to supervised image classification. VP trees were found to be superior for finding similar grey-value patches in a recent comparison against ball trees, kd-trees and hierarchical k-means [10]. We will present **vantage point forests** as new classifier for binary strings in Sec. 2.2. The advantage compared to random decision trees is that at each node the path of the sample(s) is dependent on the full feature vector and not only a single (optimised) feature dimension as in [7]. Therefore high-dimensional hyperspheres can efficiently partition the (potentially sparsely populated) feature space. Oblique decision trees [11, 12] also use multiple feature dimensions to create hyperplanes for splitting, but their training procedure is much more time-consuming (akin to ball trees cf. [10]) than our vantage point approach. We applied the proposed fully-automatic segmentation algorithm to clinical CT scans of the abdomen and demonstrate significantly improved accuracy compared to random decision forests in Sec. 3.

2 Methods

We perform patch-based classification, where a label $y_i \in \{0, 1, \ldots, |C| - 1\}$ should be assigned out of a set of classes C to each test sample i. Each image patch $\mathcal{P}_i \in \mathbb{R}^{|L|}$ (with $|L|$ pixels) associated with i can be described by a feature vector $\mathbf{h}_i \in \mathbb{H}^n$, which resides (without loss of generality) in an n-dimensional



a) concept of pair-wise intensity comparisons b) BRIEF- and LBP-like comparisons c) Contextual similarity in other scan

Fig. 1: Contextual information (BRIEF) is captured by comparing mean values of two offset locations $\mathcal{P}_i(q)$ and $\mathcal{P}_i(r)$. Structural content (LBP) can be obtained by fixing one voxel to be the central $\mathcal{P}_i(0)$. When determining the training samples from c) that are closest to the central voxel in a) using our vantage point forest the similarity map overlaid to c) is obtained, which clearly outlines the corresponding psoas muscle.

Hamming space, where $h_{id} \in \{\pm 1\}$ describes the d-th dimension of sample i. In a supervised training stage a ground truth class label has been assigned to a large number of feature vectors yielding $|M|$ training samples $(\mathbf{h}_j, y_j)$, $j \in M$. During testing a probability distribution $p(y|\mathbf{h}_i)$ is estimated for each test sample i and the most likely label $y_i^* = \text{argmax}_{y \in C}\, p(y|\mathbf{h}_i)$ is chosen.

2.1 Contextual Binary Similarity

We employ a strong combination of numerous weak intensity comparison features. In [4] contextual features, related to local binary patterns (LBP) [13] have been used to localise organs. For each binary feature the mean intensity of a region around the voxel of interest is compared to a region with a certain spatial offset. However, by relying strongly on relations of the central region, helpful pairwise interactions of two different neighbouring structures may be missed. [6] showed that by using two different offsets for both regions (i.e. none is centred at the voxel of interest) keypoints recognition rates can be much improved (using the binary BRIEF descriptor). For any given intensity patch $\mathcal{P}_i$ the feature values are simply obtained as $h_{id} = +1$ if $\mathcal{P}_i(q) > \mathcal{P}_i(r)$ for $(q, r) \in L$ and $h_{id} = -1$ else. The patch may be smoothed prior to the pixel comparisons by a Gaussian kernel with variance σ_p^2.

In Fig. 1 the proposed combination of LBP- and BRIEF-like features is visualised. Similar features based on the value of intensity differences have also been used in [1] for anatomy localisation and segmentation. We use only the sign of these differences, which improves robustness against contrast variations [4] and reduces in this work the computational complexity of the similarity calculation between two very long descriptors (by using `popcount` instructions).

2.2 Vantage Point Forests

Different approaches could be employed to determine the class probability $p(y|\mathbf{h}_i)$ for an unseen test sample i. For random forest ensembles a number of uncorrelated decision trees is trained, in which each node determines whether a sample should be directed to the left or right branch based on a selected feature dimension d and a threshold τ. Both feature dimension and threshold can be optimised during training in order to divide samples of different classes (and thus decrease the entropy of class histograms in the following levels).

Our proposed vantage point classifier, in contrast, first randomly selects a sample j out of the data available at the current node and finds a distance threshold τ so that approximately half the samples are closer to j than τ and the other half is farther away. In our work the distance between samples is defined by the Hamming weight of their entire feature vectors, so that all i for which $d_H(i,j) = ||\mathbf{h}_i - \mathbf{h}_j||_{\mathbb{H}} < \tau$ will be assigned to the left node (and vice-versa). The Hamming distance measures the number of differing bits in two binary strings $||\mathbf{h}_i - \mathbf{h}_j||_{\mathbb{H}} = \Xi\{\mathbf{h}_i \oplus \mathbf{h}_j\}$, where $\oplus$ an exclusive OR and Ξ a bit count. The partitioning is recursively repeated until a minimum leaf size is reached (we store both the class distribution and the indices of the remaining training samples S_l for each leaf node l).

During testing each sample (query) is inserted in a tree starting at the root node. Its distance w.r.t. the training sample of the current vantage point is calculated and compared with τ (determining the direction the search branches off). When reaching a leaf node the class distribution is retrieved and averaged across all trees within the forest[1].

Algorithm 1: Training of Vantage Point Forest

Input: $|M|$ labelled training samples $(\mathbf{h}_j, y_j)$, **parameters:** number of trees T, minimum leaf size $\mathcal{L}_{\min}$

Output: T tree structures: indices of vantage points, thresholds τ for every node, class distributions $p(y|\mathbf{h}_i)$ and sample indices for leaf nodes.

- **foreach** $t \in T$ **do**
  - add initial subset $S_0 = M$ (whole training set→root) to top of stack
  - **while** *stack is not empty* **do**
    - retrieve S_n from stack, select vantage point $j \in S_n$ (randomly)
    - **if** $|S_n| > \mathcal{L}_{\min}$ **then**
      - calculate $d_H(i,j) = ||\mathbf{h}_i - \mathbf{h}_j||_{\mathbb{H}} \forall i \in S_n$, and median distance $\tau = \tilde{d}_H$
      - partition elements i of S_n in two disjunct subsets $S_{nl} = \{i | d_H(i,j) < \tau\}$, $S_{nr} = S_n \setminus S_{nl}$ and add them to stack
    - **else**
      - store $p(y|\mathbf{h}_i)$ and sample indices of S_l (leaf node)

[1] Our source code is publicly available at `http://mpheinrich.de/software.html`

When trees are not fully grown (leaving more than one sample in each leaf node), we propose to gather all training samples from all trees that fall in the same leaf node (at least once) and perform a linear search in Hamming space to determine the k-nearest neighbours (this will be later denoted as **VPF+kNN**). Even though intuitively this will add computationally cost, since more Hamming distances have to be evaluated, this approach is faster in practice (for small $\mathcal{L}_{\min}$) compared to deeper trees due to cache efficiencies. It is also much more efficient than performing an approximate global nearest neighbour search using locality sensitive hashing or related approaches [14].

Split Optimisation: While vantage point forests can be built completely unsupervised, we also investigate the influence of supervised split optimisation. In this case the vantage points are not fully randomly chosen (as noted in Line 4 of Alg. 1), but a small random set is evaluated based on the respective information gain (see [7] for details on this criterion) and the point that separates classes best, setting τ again to the median distance for balanced trees, is chosen.

2.3 Spatial Regularisation using Multi-Label Random Walk

Even though the employed features provide good contextual information, the classification output is not necessarily spatially consistent. It may therefore be beneficial for a dense segmentation task to spatially regularise the obtained probability maps $P^y(\mathbf{x})$ (in practice the classification is performed on a coarser grid, so probabilities are first linearly interpolated). We employ the multi-label random walk [15] to obtain a smooth probability map $P(\mathbf{x})^y_{reg}$ for every label $y \in C$ by minimising $E(P(\mathbf{x})^y_{reg})$:

$$\sum_{\mathbf{x}} \frac{1}{2}(P(\mathbf{x})^y - P(\mathbf{x})^y_{reg})^2 + \sum_{\mathbf{x}} \frac{\lambda}{2} ||\nabla P(\mathbf{x})^y_{reg}||^2 \qquad (1)$$

where the regularisation weight is λ. The gradient of the probability map is weighted by $w_j = \exp(-(I(\mathbf{x}_i) - I(\mathbf{x}_j))^2/(2\sigma_w^2))$ based on differences of image intensities I of $\mathbf{x}_i$ and its neighbouring voxels $\mathbf{x}_j \in \mathcal{N}_i$ in order to preserve edges. Alternatively, other optimisation techniques such as graph cuts or conditional random fields (CRF) could be used, but we found that random walk provided good results and low computation times.

3 Experiments

We performed automatic multi-organ segmentations for 20 abdominal contrast enhanced CT scans from the VISCERAL Anatomy 3 training dataset (and additionally for the 10 ceCT test scans) [16]. The scans form a heterogenous dataset with various topological changes between patients. We resample the volumes to 1.5 mm isotropic resolution. Manual segmentations are available for a number of different anatomical structures and we focus on the ones which are most frequent in the dataset, namely: liver, spleen, bladder, kidneys and psoas major muscles (see example in Fig. 2 with median automatic segmentation quality).

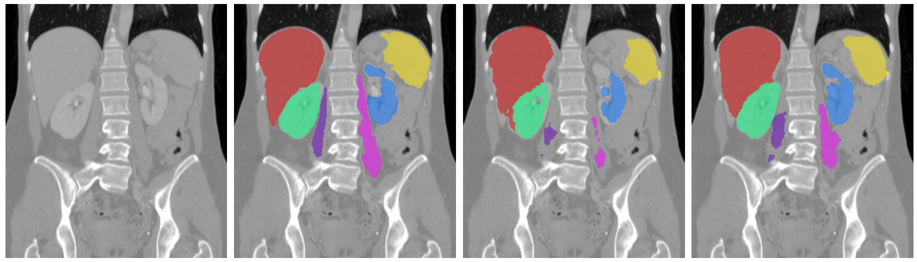


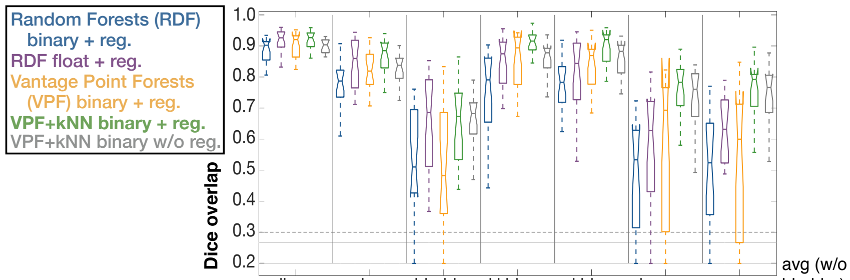
Fig. 2: Coronal view of CT segmentation: Psoas muscles ■■ and left kidney ■ are not fully segmented using random forests. Vantage point forests better delineate the spleen ■ and the interface between liver ■ and right kidney ■ (bladder is out of view).

Parameters: Classification is performed in a leave-one-out fashion. A rough foreground mask (with approx. 30 mm margin to any organ) is obtained by nonrigidly registering a mean intensity template to the unseen scan using [17]. We compare our new vantage point classifier to standard random forests (**RDF**) with axis-aligned splits using the implementation of [18]. For each method 15 trees are trained and either fully grown or terminated at a fixed leaf size of $\mathcal{L}_{\min} = 15$ (**VPF+kNN**). Using more trees did not improve classification results of **RDF**. The number k of nearest neighbours in **VPF+kNN** is set to 21.

A total of $n = 640$ intensity comparisons are used for all methods within patches of sizes of 101^3 voxels, after pre-smoothing the images with a Gaussian kernel with $\sigma_p = 3$ voxels. Half the features are comparisons between the voxel centred around i and a randomly displaced location (LBP), and for the other half both locations are random (BRIEF). The displacement distribution is normal with a standard deviation of 20 or 40 voxels (for 320 features each). The descriptors are extracted for every fourth voxel (for testing) or sixth voxel (in training) in each dimension (except outside the foreground mask) yielding ≈500'000 training and ≈60'000 test samples. Spatial regularisation (see Sec. 2.3) is performed for all methods with optimal parameters of $\lambda = 10$ for **RDF**, $\lambda = 20$ for **VPF** and $\sigma_w = 10$ throughout (run time ≈ 20 sec.).

RDF have been applied with either binary or real-valued (float) features. We experimented with split-node optimisation for **VPF**, but found (similar to [3] for ferns) that it is not necessary unless when using very short feature strings (which may indicate that features of same organs cluster together without supervision).

Results: We evaluated the automatic segmentation results A using the Dice overlap $D = 2|A \cap E|/(|A| + |E|)$ (compared to an expert segmentation E). Vantage point forests clearly outperform random forests and achieve accuracies of >0.90 for liver and kidneys and ≈0.70 for the smaller structures. Random forests benefit from using real-valued features but are on average 10 percentage points inferior, revealing in particular problems with the thin psoas muscles. Our average Dice score of 0.84 (see details in Fig. 3) is higher than results for MALF: 0.70 or SIFT keypoint transfer: 0.78 published by [8] on the same VISCERAL



		liver	spleen	bladder	l kidney	r kidney	l pmm	r pmm	avg (w/o bladder)
training set	**VPF+kNN binary+reg.**	**0.91**	**0.83**	**0.62**	**0.90**	**0.90**	**0.75**	**0.75**	**0.84**
	RDF float+reg	0.90	0.79	0.61	0.81	0.79	0.55	0.58	0.74
	RDF binary+reg	0.88	0.75	0.51	0.74	0.76	0.47	0.49	0.68
	Keypoint Transfer [8]	0.83	0.77		0.87	0.86	0.67	0.67	0.78
	ANTs + local MALF [8]	0.83	0.71		0.74	0.71	0.63	0.62	0.70
test set	**VPF+kNN binary+reg.**	**0.93**	**0.87**	**0.77**	**0.91**	**0.92**	**0.82**	**0.80**	**0.88**
	Wang (shape model) [16]	0.95	0.91	0.87	0.95	0.96	0.83	0.85	0.91
	Kéchichian (MALF) [16]	0.93	0.90	0.82	0.92	0.92	0.80	0.81	0.88

Fig. 3: Distribution of Dice overlaps demonstrates that vantage point forests significantly outperform random forests ($p < 0.001$) and improve over several algorithms from the literature. Including the kNN search over samples within leaf nodes from all trees is particularly valuable for the narrow psoas muscles. Our results are very stable across all organs and not over-reliant on post-processing (see boxplots with grey lines).

training set. For the test set [16], we obtain a Dice of 0.88, which is on par with the best MALF approach and only slightly inferior to the overall best performing method that uses shape models and is orders of magnitudes slower. Training times for vantage point trees are ≈15 sec. (over 6x faster than random forests). Applying the model to a new scan takes ≈1.5 sec. for each approach.

4 Conclusion

We have presented a novel classifier, vantage point forest, that is particularly well suited for multi-organ segmentation when using binary context features. It is faster to train, less prone to over-fitting and significantly more accurate than random forests (using axis-aligned splits). VP forests capture joint feature relations by comparing the entire feature vector at each node, while being computationally efficient (testing time of ≈1.5 sec.) due to the use of the Hamming distance (which greatly benefits from hardware `popcount` instructions, but if necessary real-valued features could also be employed in addition). We demonstrate state-of-the-art performance for abdominal CT segmentation – comparable to much more time-extensive multi-atlas registration (with label fusion). We obtained especially good results for small and challenging structures. Our method would also be directly applicable to other anatomies or modalities such as MRI, where the contrast insensitivity of BRIEF features would be desirable. The results of our algorithm could further be refined by adding subsequent stages (cascaded classification) and be further validated on newer benchmarks e.g. [19].

References

1. Glocker, B., Pauly, O., Konukoglu, E., Criminisi, A.: Joint classification-regression forests for spatially structured multi-object segmentation. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., eds.: ECCV 2012. LNCS, vol. 7575. Springer, Heidelberg (2012) 870–881
2. Breiman, L.: Random forests. Machine learning **45**(1) (2001) 5–32
3. Özuysal, M., Calonder, M., Lepetit, V., Fua, P.: Fast keypoint recognition using random ferns. IEEE PAMI **32**(3) (2010) 448–461
4. 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 Peters, T., Fichtinger, G., Martel, A., eds.: MICCAI 2011. LNCS, vol. 6893. Springer, Heidelberg (2011) 239–247
5. Zikic, D., Glocker, B., Criminisi, A.: Encoding atlases by randomized classification forests for efficient multi-atlas label propagation. Medical Image Analysis **18**(8) (2014) 1262–1273
6. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: Computing a local binary descriptor very fast. IEEE PAMI **34**(7) (2012) 1281–1298
7. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends. Comp. Graphics and Vision **7**(2–3) (2012) 81–227
8. Wachinger, C., Toews, M., Langs, G., Wells, W., Golland, P.: Keypoint transfer segmentation. In Ourselin, S., Alexander, D., Westin, C., Cardoso, M., eds.: IPMI 2015. LNCS, vol. 9123, Springer, Heidelberg (2015) 233–245
9. Yianilos, P.N.: Data structures and algorithms for nearest neighbor search in general metric spaces. In: SODA. Volume 93. (1993) 311–321
10. Kumar, N., Zhang, L., Nayar, S.: What is a good nearest neighbors algorithm for finding similar patches in images? In Forsyth, D., Torr, P., Zisserman, A., eds.: ECCV 2008. LNCS, vol. 5303. Springer, Heidelberg (2008) 364–378
11. Menze, B.H., Kelm, B.M., Splitthoff, D.N., Koethe, U., Hamprecht, F.A.: On oblique random forests. In: ECML. (2011) 453–469
12. Schneider, M., Hirsch, S., Weber, B., Székely, G., Menze, B.: Joint 3-d vessel segmentation and centerline extraction using oblique hough forests with steerable filters. Medical Image Analysis **19**(1) (2015) 220–249
13. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE PAMI **28**(12) (2006) 2037–2041
14. Muja, M., Lowe, D.G.: Fast matching of binary features. In: CRV. (2012) 404–410
15. Grady, L.: Multilabel random walker image segmentation using prior models. In: CVPR. (2005) 763–770
16. Jiménez-del Toro, O., et al.: Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks. IEEE Trans Med Imag (2016) 1–20
17. Heinrich, M., Jenkinson, M., Brady, J., Schnabel, J.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans Med Imag **32**(7) (2013) 1239–1248
18. Dollar, P., Rabaud, V.: Piotr Dollar's image and video toolbox for matlab. UC San Diego (2013) available at: `https://github.com/pdollar/toolbox`.
19. Xu, Z., Lee, C., Heinrich, M., Modat, M., Rueckert, D., Ourselin, S., Abramson, R., Landman, B.: Evaluation of six registration methods for the human abdomen on clinically acquired CT. IEEE Trans Biomed Eng (2016) 1–10