Geodesic Patch-Based Segmentation

Zehan Wang[1], Kanwal K. Bhatia[1], Ben Glocker[1], Antonio Marvao[2], Tim Dawes[2], Kazunari Misawa[3], Kensaku Mori[4,5], and Daniel Rueckert[1]

[1] Biomedical Image Analysis Group, Department of Computing, Imperial College London, London, UK
[2] Institute of Clinical Sciences, Imperial College London, London, UK
[3] Aichi Cancer Center, Nagoya, Japan
[4] Department of Media Science, Nagoya University, Nagoya, Japan
[5] Information and Communications Headquarters, Nagoya University, Nagoya, Japan

Abstract. Label propagation has been shown to be effective in many automatic segmentation applications. However, its reliance on accurate image alignment means that segmentation results can be affected by any registration errors which occur. Patch-based methods relax this dependence by avoiding explicit one-to-one correspondence assumptions between images but are still limited by the search window size. Too small, and it does not account for enough registration error; too big, and it becomes more likely to select incorrect patches of similar appearance for label fusion. This paper presents a novel patch-based label propagation approach which uses relative geodesic distances to define patient-specific coordinate systems as spatial context to overcome this problem. The approach is evaluated on multi-organ segmentation of 20 cardiac MR images and 100 abdominal CT images, demonstrating competitive results.

1 Introduction

Accurate segmentation in medical imaging plays a crucial role in many applications from patient-specific diagnosis to population studies. The ability to perform this task without human intervention is particularly desirable for large datasets. Multi-atlas label propagation approaches [7],[10],[1], in which labels from multiple atlases are propagated to target images after registration, have been shown to be highly effective. However, dependence on image registration for these methods can be problematic as inaccurate alignment adversely affects segmentation quality. Additionally, finding suitable (fixed) registration parameters that yield accurate non-linear correspondences on different images can be a challenge on its own, particularly for anatomies that are highly variable. Patch-based methods for label propagation [3],[11] can help alleviate this dependence since they do not rely on explicit one-to-one correspondences between images, and are often able to use affine rather than non-rigid registration, yet still produce comparable results.

Patch-based segmentation assumes that patches with similar intensities and from similar local neighbourhoods are likely to be the part of the same anatomical structure. Traditionally, this locality is enforced by a sliding search window

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 666–673, 2014.

of a fixed size (typically $< 11^3$ voxels). Label fusion then determines spatially-varying weights for each label according to the similarity of the corresponding patches within each voxel neighbourhood. This neighbourhood, when defined as a fixed size search window, imposes a hard restriction on tolerance to any registration errors that occur. Increasing the size increases the tolerance to registration errors but also increases the computational requirements and may yield patches with similar appearance but from different anatomical structures. Using hierarchical frameworks [5], [15] partly addresses these restrictions, however these approaches still use a fixed search window size. More recently, several methods have reformulated the standard patch-based approach to consider the local neighbourhood for each voxel globally based on k nearest neighbours (kNN) [13] or a trained neighbourhood approximation [8] rather than using a fixed search window size. This alleviates the computational burden, whilst the selection of patches are regularised by different approaches to apply spatial context and consistency. Another approach uses random forests trained on individual atlases with spatially varying representation of patches instead of fixed patch sizes [16].

The use of spatial context regularises the patch selection process by comparing spatial similarities between patches as well as their appearance similarity. This enables locally similar patches of different structures to be distinguished when larger search windows are used and also increases the tolerance with regards to variability between images in intensities for the same structures. One approach to incorporating spatial context is to use spatial coordinates and euclidean distances between labelled structures [13], [6]. However, this does not take into account the context of the image such as the boundaries between structures and is sensitive to anatomical variability when comparing between subjects. In this paper, we propose instead to use geodesic distances within the image to provide spatial context which is able to contribute information on the locality of structure boundaries. The use of geodesic distances has been shown to be effective in interactive segmentation [4] and we adopt this within an automated patch-based segmentation method, formulating a multi-resolution approach based on adaptive, anatomically-specific coordinates in order to leverage its use. We implement this within a kNN framework using fast-building kNN data structures, so that these adaptive spatial features can be applied at run time. We evaluate our proposed methods on multi-structure and multi-organ segmentation of 20 cardiac MR images and 100 abdominal CT images, respectively.

2 Methods

Our framework is intended to extend the ability of patch-based segmentation methods to tolerate potential registration errors. Increasing this tolerance requires patch comparisons to be made within a search space that can encompass the margin of error but also maintain the sense of "locality" which restricts the comparison to relevant patches. To do this, we define the local neighbourhood for each patch by its kNN in terms of both spatial context and intensity similarity, so as to distinguish between similar patches from different structures. This

allows the search space of patches to be global whilst maintaining the sense of locality, thus removing the requirement for a fixed search window size to be set.

2.1 Adaptive Coordinate System as Spatial Context

When dealing with potentially large misalignments between images, spatial context based on explicit image coordinates can be unreliable. Spatial context should therefore be defined in a way that is robust to misregistrations. To this end, we introduce the concept of *relative distances*. If we can establish any reference points or an initial rough segmentation, we can then use a distance transform on the labelled structures to create a non-Cartesian, *patient-specific coordinate system* that is invariant to how anatomical structures are positioned within the image (Fig. 1a, 1b). For a voxel x, we define the spatial context $S(x)$ as a vector $[d(x, r_1), d(x, r_2), ..., d(x, r_n)]$ where $d(.)$ is the distance and $r_1, ..., r_n$ are labelled structures or landmarks.

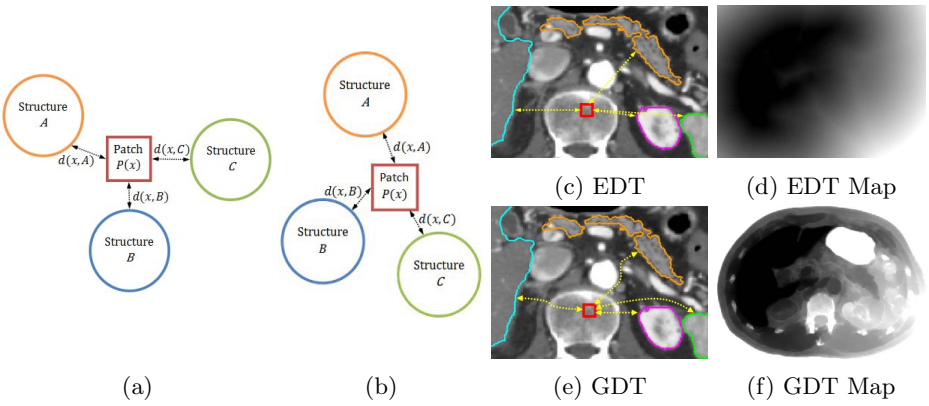


Fig. 1. (a) and (b) shows how spatial context for x can be provided by the distances to structures regardless of how they are positioned within the image. (c) and (e) provide an example where this type of spatial context can be used and how the distances will be different. (d) and (f) show the respective distance maps using EDT and GDT from the liver, where distances are shown as being proportional to the intensity.

In principle, at least three reference points are required to localise a point in 3D space, but useful information can still be obtained with less. For example, relative distances from two structures can localise a curve, whilst distances from a single structure can provide enough spatial context for a surface. This may be enough to distinguish between patches of similar intensities from different structures.

One possible approach for relative distances is to use the Euclidean Distance Transform (EDT), however the EDT does not take information present within the image into account, such as visible boundaries between structures, and may

be insufficient when high anatomical variation exists (see Fig. 1). To improve localisation, we propose to additionally include information from visible boundaries, given by image gradients, between structures.

2.2 Geodesic Distance Transform (GDT)

To overcome the shortfalls of the EDT in providing adaptive spatial context, we propose the use of a geodesic distance transform (GDT), which takes into account image gradients and describes distances between structures using the shortest path along the image intensities rather than just through physical (empty) space [4]. In general, the geodesic distance between two points x, y within an image I is defined as follows:

$$d(x,y) = \inf_{\mathbf{\Gamma} \in \mathbf{P}_{x,y}} \int_0^{l(\mathbf{\Gamma})} \sqrt{1 + \gamma^2 (\nabla I(s) \cdot \mathbf{\Gamma}'(s))^2} \mathrm{d}s \tag{1}$$

where $\mathbf{\Gamma}$ is a path in the set of all paths, $\mathbf{P}_{x,y}$ between x and y and is parametrised by its arclength $s \in [0, l(\mathbf{\Gamma})]$. The EDT can be considered a special case of the GDT, since these are equivalent when γ is set to 0. To calculate the GDT, we use the approach from [12] which was also used in [4] and demonstrated to have good performance with linear computational requirements.

2.3 Spatially Weighted Label Fusion

With the inclusion of spatial context, the label fusion for voxel x is determined as follows - let $P(x)$ be a vector of intensities for the patch at voxel x and let $\{y_{L,i} : i \in 1, ..., k\}$ represent potential matches from the atlas library for each label L. A weighting w_L for each label L is determined by comparing the k nearest patches from the atlas library with regards to both intensities $P(x)$ and spatial context $S(x)$ - and as such no search window is used:

$$w_L(x) = \sum_{i=1}^{k} w(x, y_{L,i}) \tag{2}$$

where

$$w(x,y) = e \frac{-\{||P(x) - P(y)||_2^2 + \alpha ||S(x) - S(y)||_2^2\}}{h^2(x)} \tag{3}$$

and similarly to [3], $h^2(x)$ is determined by the minimum distance:

$$h^2(x) = \min\{||P(x) - P(y_i)||_2^2 + \alpha ||S(x) - S(y_i)||_2^2\} \tag{4}$$

A spatial weighting α balances the relative importance between the spatial and intensity components. The final label $\hat{L}$ at voxel x is decided by majority vote, i.e. $\hat{L}(x) = \arg\max_L w_L(x)$.

2.4 Framework Implementation

As an overall segmentation framework, we propose applying a multi-resolution approach with the core methods described above in an iterative process where only the boundaries, defined by the difference between the dilation and erosion of each segmentation, are refined as the segmentation is propagated through higher resolutions [13]. Multiple resolutions of each image can be created offline by constructing a Gaussian image pyramid. This reduces the computational cost compared to processing directly at the native resolution.

The GDT is always calculated in the native resolution (and downsampled if required) so that the same spatial weighting α can be used at all resolutions. There are several options for an initial segmentation to enable the use of the GDT for spatial context. For instance, the intersections of the atlases can be used if this does not yield an empty set. However this may not always occur, in which case, a coarse segmentation can be established in the lowest resolution using coordinates as spatial context [14] or another segmentation technique. These can be eroded, and relative distances can be calculated from eroded versions of each structure to reduce the initial error. Successive refinements of the boundary regions reduces the dependence on the initial segmentation.

For patch selection, it is highly desirable to have an efficient kNN data structure, since performing global kNN search can be a computational bottleneck. In our implementation, ball trees were used since they (and metric trees in general) have been shown to have better performance in higher dimensional spaces (> 20) compared to space-partitioning structures like the kd-tree [9].

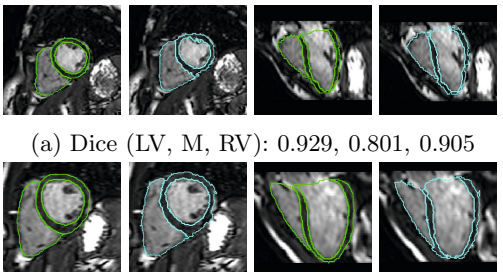
3 Experiments and Results

3.1 Cardiac MR Dataset

We applied our proposed approach and compared it to using voxel coordinates and the EDT as spatial context, in addition to the standard patch-based approach from [3] (Coupé) in segmenting the left ventricle, myocardium and right ventricle in end diastole frames of 20 subjects under a single breath-hold. The MR images were captured from a 1.5T Philips Achieva system and have a native resolution of $256\times256\times64$ voxels with voxel sizes of $1.25\times1.25\times2\text{mm}^3$.

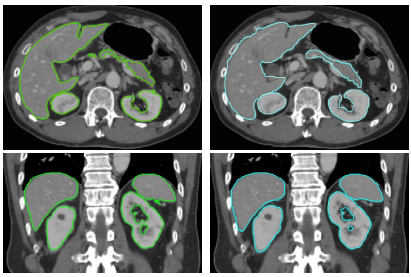
In total, 3 resolution levels were used by our approach, with the lowest resolution at $5\times5\times5\text{mm}^3$ voxel sizes and the intermediate level at $2.5\times2.5\times2.5\text{mm}^3$ voxel sizes. Images were aligned using affine registration with 6 manually placed landmarks, and the intersections of the atlases were used as initial segmentations. We evaluated each method using leave-one-out cross validation and used the all available atlases (19) to segment each test image. A patch size of $5\times5\times5$ voxels was used for all resolutions and k was fixed at 40 for the kNN methods. For the different approaches to spatial context, α was selected in the lowest resolution and then applied for all subsequent resolutions whilst γ was set at 100 for GDT. The values for k and γ were not tuned.

Table 1 and Fig. 3 summarises the final segmentation accuracy and examples of segmented images using our proposed method are presented in Fig. 2.



(a) Dice (LV, M, RV): 0.929, 0.801, 0.905

(b) Dice (LV, M, RV): 0.953, 0.838, 0.934



(c) Dice (L, S, P, K): 0.947, 0.955, 0.826, 0.924

Fig. 2. Examples from using our approach (in cyan) for the Left Ventricle (LV), Myocardium (M) and Right Ventricle (RV) in (a) and (b), and Liver (L), Spleen (S), Pancreas (P), Kidneys (K). Reference segmentations are outlined in green.

Table 1. Overall Dice Coefficients shown as mean (median) ±standard deviation

Method/Description	Left Ventricle	Myocardium	Right Ventricle
Coordinates $\alpha = 7$	0.931 (0.934) ±0.016	0.763 (0.763) ±0.049	0.871 (0.879) ±0.037
EDT $\alpha = 13$	0.938 (0.938) ±0.017	0.806 (0.814) ±0.049	0.882 (0.893) ±0.047
GDT $\alpha = 5$	0.934 (0.941) ±0.019	0.797 (0.803) ±0.039	0.901 (0.904) ±0.021
[3] window size = 7^3	0.931 (0.936) ±0.020	0.773 (0.787) ±0.053	0.889 (0.902) ±0.035

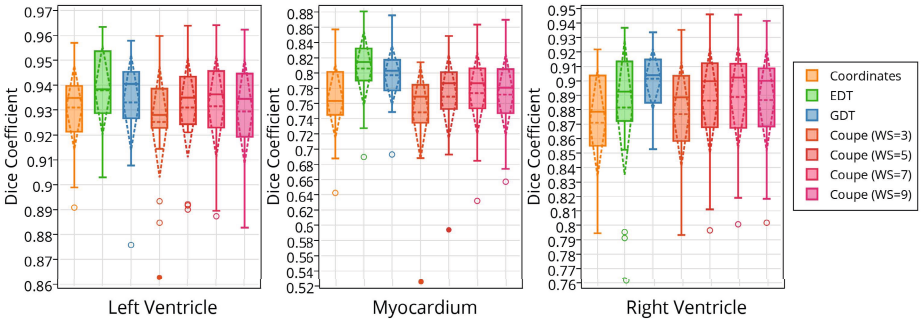


Fig. 3. Comparison of segmentation accuracy with regards to different spatial context and different search window sizes (WS=x). Solid line represents the median, the dashed lines represents the mean and standard deviation.

3.2 Abdominal CT Dataset

We also evaluated our method on 100 abdominal CT scans which have an in plane resolution of 512×512 voxels with voxel sizes ranging from 0.55 to 0.82mm and contain between 263 to 538 slices with spacing ranging from 0.4 to 0.8mm depending on the field of view and the slice thickness. For each scan, manual segmentations of the liver, spleen, pancreas and the kidneys were generated by a single trained rater.

Table 2. Overall Dice Coefficients shown as mean±standard dev. [worst, best]

Organ	Ours	[15]
Liver	0.945±0.025 [0.842, 0.977]	0.940±0.028 [0.814, 0.974]
Spleen	0.925±0.084 [0.461, 0.979]	0.920±0.092 [0.264, 0.982]
Pancreas	0.655±0.186 [0.024, 0.902]	0.696±0.167 [0.069, 0.909]
Kidneys	0.924±0.077 [0.334, 0.982]	0.925±0.072 [0.515, 0.982]

For our approach, four resolutions levels were used, ranging from 4mm^4 to 1mm^3 voxel sizes. For each test image, affine registration (without landmarks) was used to align the atlases, with the 50 nearest chosen using sum of squared differences as the distance measure. Initial coarse segmentations were established by using coordinates as spatial context [14] with $\alpha = 3$, whilst subsequent refinements using GDT as spatial context used $\alpha = 7$. Our results are presented in Table 2 alongside those from [15] where affine registration followed by additional organ level non-rigid deformable registrations were used. Similarly to [15], we apply graph cuts [2] as post processing to obtain the final segmentation.

4 Discussion and Conclusion

This paper has presented a new patch-based segmentation approach which uses spatial context to provide robustness to misregistrations. To do this, we defined an adaptive, anatomically-specific coordinate system based on relative distances between structures and used geodesic distances to be able to localise even highly variable anatomy. Our approach produced results which are competitive with existing patch-based methods. The ability to do so on abdominal data with only affine registration when previous methods have also used non-rigid deformable registration demonstrates the robustness of our approach. Our implementation took around 4 hours to segment each cardiac image (with 19 atlases) and 14 hours for each abdominal image (with 50 atlases) using 16 CPU cores clocked at 2.8Ghz. However, this can be significantly reduced with a more optimal kNN data structure. Also, our framework is easily parallelisable, since voxels are labelled independently, and is very scalable with parallel hardware. Furthermore, kNN patch search for each atlas is performed independently, enabling usage of distributed networks. Overall this approach shows much potential, particularly in more challenging datasets where achieving accurate registration is difficult.

References

1. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46(3), 726–738 (2009)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. IEEE PAMI 23(11), 1222–1239 (2001)

3. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage 54(2), 940–954 (2011)
4. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic Image Segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
5. Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Østergaard, L.R., Collins, D.L.: BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59(3), 2362–2373 (2012)
6. 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, Part IV. LNCS, vol. 7575, pp. 870–881. Springer, Heidelberg (2012)
7. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115–126 (2006)
8. Konukoglu, E., Glocker, B., Zikic, D., Criminisi, A.: Neighbourhood approximation using randomized forests. Medical Image Analysis 17(7), 790–804 (2013)
9. Kumar, N., Zhang, L., Nayar, S.K.: What Is a Good Nearest Neighbors Algorithm for Finding Similar Patches in Images? In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 364–378. Springer, Heidelberg (2008)
10. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.R.: Evaluation of Atlas Selection Strategies for Atlas-Based Image Segmentation with Application to Confocal Microscopy Images of Bee Brains. NeuroImage 21(4), 1428–1442 (2004)
11. Rousseau, F., Habas, P., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. IEEE TMI 30(10), 1852–1862 (2011)
12. Toivanen, P.J.: New geodosic distance transforms for gray-scale images. Pattern Recognition Letters 17(5), 437–450 (1996)
13. Wang, Z., Donoghue, C., Rueckert, D.: Patch-based segmentation without registration: Application to knee MRI. In: Wu, G., Zhang, D., Shen, D., Yan, P., Suzuki, K., Wang, F. (eds.) MLMI 2013. LNCS, vol. 8184, pp. 98–105. Springer, Heidelberg (2013)
14. Wang, Z., Wolz, R., Tong, T., Rueckert, D.: Spatially Aware Patch-Based Segmentation (SAPS): An Alternative Patch-Based Segmentation Framework. In: Menze, B.H., Langs, G., Lu, L., Montillo, A., Tu, Z., Criminisi, A. (eds.) MCV 2012. LNCS, vol. 7766, pp. 93–103. Springer, Heidelberg (2013)
15. Wolz, R., Chu, C., Misawa, K., Fujiwara, M., Mori, K., Rueckert, D.: Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE TMI 32(9), 1723–1730 (2013)
16. Zikic, D., Glocker, B., Criminisi, A.: Atlas Encoding by Randomized Forests for Efficient Label Propagation. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part III. LNCS, vol. 8151, pp. 66–73. Springer, Heidelberg (2013)