Patch-based Segmentation without Registration: Application to Knee MRI

Zehan Wang, Claire Donoghue, Daniel Rueckert

Department of Computing, Imperial College London, UK
zehan.wang06@imperial.ac.uk

Abstract. Atlas based segmentation techniques have been proven to be effective in many automatic segmentation applications. However, the reliance on image correspondence means that the segmentation results can be affected by any registration errors which occur, particularly if there is a high degree of anatomical variability. This paper presents a novel multi-resolution patch-based segmentation framework which is able to work on images without requiring registration. Additionally, an image similarity metric using 3D histograms of oriented gradients is proposed to enable atlas selection in this context. We applied the proposed approach to segment MR images of the knee from the MICCAI SKI10 Grand Challenge, where 100 training atlases are provided and evaluation is conducted on 50 unseen test images. The proposed method achieved good scores overall and is comparable to the top entries in the challenge for cartilage segmentation, demonstrating good performance when comparing against state-of-the-art approaches customised to Knee MRI.

1 Introduction

Accurate segmentation in medical imaging plays a crucial role in many applications from patient-specific diagnosis to biomedical research. Patch-based methods for label propagation have been shown to be an effective automatic segmentation approach for many applications in medical imaging [4], [14]. In general, these approaches label each voxel of a target image by comparing the image patch centred on the voxel with patches from an atlas library and assigning the most probable label according to the closest matches. Most of these methods use affine registration rather than non-rigid registration which is often used in multi-atlas label propagation methods [8], [13], yet are still able to produce comparable results.

One advantage of patch-based approaches is that they do not assume explicit one-to-one correspondence between images, which can overcome problems for multi-atlas segmentation where the anatomical variability cannot be fully accounted for when generalised registration algorithms are used with the available atlases. However, due to computational complexity, many existing patch-based methods use limited search windows centred around the target voxel (typically around 11^3 voxels) and thus rely upon a good approximate alignment of the image to the atlases. Increasing the search window size also increases the number of

patch comparisons required, which can be computationally prohibitive, limiting the potential applications of this approach. Hierarchical frameworks have been previously used to address these restrictions, such as multi-resolution solutions [6] or by combining hierarchical multi-atlas registration and patch-based segmentation [19]. Another approach uses efficient k-nearest neighbour data structures and a spatially weighted label fusion method [18].

Despite these improvements, the proposed methods still require the registration of images to establish image correspondences, albeit with affine registration rather than non-rigid. This process can still be a computationally expensive and problematic step in images with a high degree of anatomical variance and where a common template space (such as MNI space for brain images) isn't defined. Problems with affine registration of knee MRI have been previously quantified in [5] where the authors observed 4.08% of direct pairwise registrations fail without manual input in a data set of over $10,000$ knee images.

To avoid these issues, we investigate the potential to perform segmentation without registration. Our first contribution is a multi-resolution framework for patch-based segmentation which does not require image registration. This approach can quickly obtain a coarse initial segmentation in the lowest resolution which is then refined by propagating through subsequently higher resolutions. Our second contribution is an atlas selection method using a histogram approach with 3D oriented gradients to provide a fast and generic similarity measure in this context where image correspondence cannot be assumed. This is a new approach for atlas selection, but in the context of image retrieval similar approaches have been developed. We refer to [1] for a review of other potential approaches that could be utilised such as those using texture, contours or wavelets.

Our third contribution is to apply the proposed methods to knee MR images from the MICCAI SKI10 Grand Challenge [9], where 100 training atlases are provided and evaluation is conducted on 50 unseen test images. There is a high degree of anatomical variability in the cartilage and bone structures in the knee which is the primary reason why segmentation of cartilage is particularly challenging. This is the first time that a patch-based method has been developed for segmentation without requiring prior registration in addition to its application to knee images. In terms of the application to the knee MRI, the proposed method can label both bones and cartilage simultaneously, whereas many other methods require a bone segmentation first to find the bone cartilage interface as an intermediate step prior to cartilage segmentation [7].

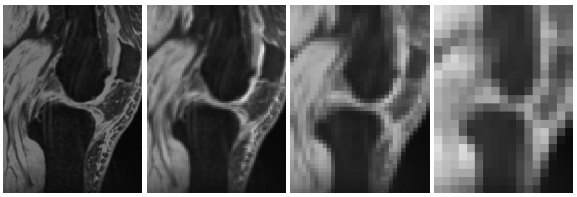
2 Methods

2.1 Hierarchical Segmentation Strategy

A multi-resolution approach allows a coarse initial segmentation to be quickly established from the lowest resolution image which can then be refined using subsequently higher resolutions. The basic concept for our patch-based approach is a simple iterative process of atlas selection, patch selection, and then label

fusion. Multiple resolutions of each image are created by constructing a Gaussian image pyramid [3]. As a pre-processing step, MR images are bias field corrected [17], re-sampled to isotropic voxel size and intensity normalised [12]. The lowest resolution is based on the findings in [16], where, motivated by the remarkable tolerance of the human visual system to degradations in image resolution, the authors reported automated segmentation tasks can be performed on images with resolutions as small as in the range of 32×32 to 16×16 depending on the size of the object. With this in mind, the sub-sampling of each image terminates if the resolution is less than $32 \times 32 \times 32$. Our initial experiments confirmed that a coarse segmentation can be performed in the lowest resolution and that a patch-based approach can be used to do so, even when the images are not registered.

Table 1. Example of a Gaussian image pyramid for an image in the SKI10 data set. Resolution levels are numbered upwards in ascending order, where level 1 denotes the highest resolution and each subsequent level is half the resolution of the previous level.



Level	1	2	3	4
Voxel size	0.78mm	1.56mm	3.12mm	6.24mm
Resolution	$\sim 140^3$	$\sim 70^3$	$\sim 35^3$	$\sim 17^3$

Given an initial segmentation from a low resolution image, only a boundary region around each, defined by the difference between the *dilation* and *erosion* of the segmentation, will need to be refined in a higher resolution. This is because a low resolution segmentation cannot represent the boundary voxels of each structure as well as a higher resolution. However, for internal voxels of each structure, the low resolution segmentation would be a sufficient, so no further refinement would be required for these voxels. Multiple iterations of this boundary refinement can be carried out at each resolution to increase the accuracy. This forms a straight-forward and computationally-efficient strategy to process images through increasing resolutions, allowing a patch-based approach to be feasibly used in all resolutions.

2.2 Atlas Selection Using Histogram of Gradients

As a trade-off between time and accuracy, only the nearest N atlas regions are used for patch-based segmentation, since linear increases in the number of atlases yields logarithmic improvements in the segmentation [18]. Without registration, existing atlas selection methods [13], [2] using voxel-to-voxel distance

metrics cannot be used meaningfully as they implicitly assume correspondences between the images. One alternative is to compare images by their intensity histogram, however this discards local neighbourhood information which is relevant for patch comparisons. Instead, we propose using a histogram of oriented 3D gradients based on [10] which incorporates the local neighbourhood information of each voxel for gradient calculations.

Gradients are calculated using a 1D Sobel operator [-1, 0, 1] in each axis and binned into one of 20 orientations based on the faces of a icosahedron. Ensuring the orientation bins are equidistant in 3D is not trivial, known otherwise as the Thomson problem [15], but using the centre positions of faces on regular polyhedrons such as the icosahedron is one solution. Histograms are built by binning the gradients of all relevant voxels using the magnitude of the gradient as the value to bin. There are many distance measures that can be employed for histogram comparison such as the chi squared distance or the earth mover's distance, but in this paper the L_1 norm was found to be a sufficiently effective.

At the start of the segmentation, the atlas selection process uses a histogram of the whole image, but once an initial segmentation has been established, only the boundary region around each label are compared - since this is the focus for the refinement process. To increase specificity, histograms are calculated separately for each label within each region and then concatenated to produce an overall histogram for that region.

2.3 Spatial Information and k-NN Data Structures

k-NN data structures are constructed for each resolution of each atlas to increase the search window size without a detrimental impact to the search speed [18]. For each voxel a feature vector is constructed from spatial information as well as intensities of the patch for that voxel. The spatial information provides context to the intensity values of each patch and acts to regularise the patch selection during segmentation. A weighting, α, is applied to control its influence. These feature vectors are grouped by the voxel's label and a k-NN data structure is constructed for each label. This can also be applied for separate corresponding regions of each atlas rather than the whole image to allow more specificity in atlas selection. This allows a different set of atlases to be used for each region and also improves the search speed by decreasing the search space.

For Segmentation Initialisation The inclusion of spatial context is not a requirement to establish an approximate segmentation, however its inclusion can improve the accuracy. Without prior image alignment, it is still possible to generalise the relative position of a patch within the image, even with varying fields of view. Rather than use exact positioning, the Euclidean distance to $n \geq 0$ reference point(s) can be used to provide weak spatial context for each voxel. The number of reference points depends on the application as increasing this would make the spatial context stronger, but assume stronger spatial correspondence which may not be the case.

For Segmentation Refinement Once an initial segmentation is established, spatial context can then be defined relative to the segmented structures, which are invariant to their position in the image. The Euclidean Distance Transform of relevant labelled structures, which can be calculated in linear time [11], can be used to do this, as this provides the Euclidean distance to the nearest surface point for each voxel.

2.4 Patch Search and Label Fusion

Labels are calculated independently for each voxel based on euclidean distances, but in contrast to the approaches used in [4] and [14], only the k nearest patches for each possible label are used in the label fusion. Since the values providing spatial context is appended to each patch, using k-NN data structures will select the most similar patches with respect to the spatial context without any further computations. The spatial weight, α, controls the influence of the spatial context on the overall patch similarity. Let y_{l_i} represent voxels from the atlas library for label l, then the weighting for this label at voxel x is then determined by a sum of the weights, $w_l(x) = \sum_{i=1}^{k} w(x, y_{l_i})$ where

$$w(x,y) = e\frac{-\{||P(x) - P(y)||_2^2 + \alpha||S(x) - S(y)||_2^2\}}{h^2(x)} \tag{1}$$

$P(x)$ and $S(x)$ are, respectively, the patch intensities and the spatial context at voxel x. $h^2(x)$ provides an automatic estimation of a decay parameter to control the level of influence of patches as the distance increases [4]. This is performed for each voxel based on the minimum distance between patch $P(x)$ and the nearest k patches $\{P(y_i)\}$:

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

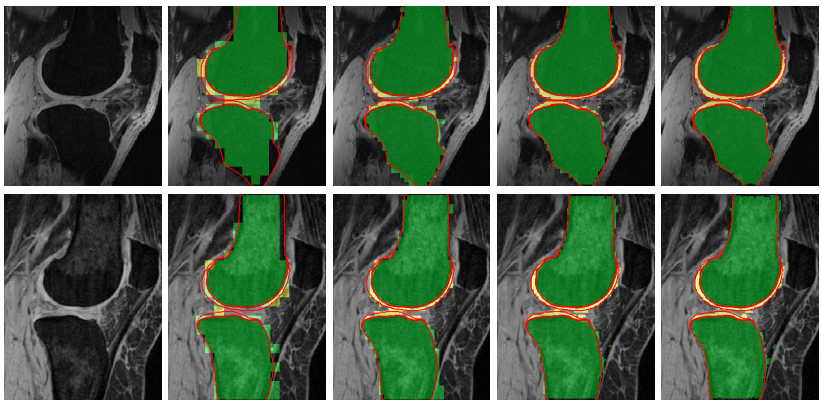
A final label for voxel x is decided based by majority voting of the label weights. i.e. $L(x) = \arg\max_l w_l(x)$

3 Application to SKI10 Grand Challenge

3.1 Implementation and Experimentation

We implemented the framework in Python with publicly available modules such as numpy, scipy and sci-kit, using a ball tree as the k-NN data structure. During experiments, the best patch sizes to use were found to be dependent on the image resolution. At the lowest resolution, a larger patch size (7^3) is used to obtain the initial coarse segmentation and smaller patch sizes (5^3 and 3^3) are then applied to refine this before propagating it to the next resolution level. This initial segmentation provides a better initialisation for the next resolution level and therefore the rest of the segmentation process.

Table 2. Example segmentations, parameters and computation times for each resolution level. Segmentations from the proposed method are overlaid in green for the bone and yellow for the cartilage. The reference segmentation is outlined in red.



Level	4	3	2	1
Patch Size	7^3, 5^3, 3^3	3^3	5^3	7^3
α	10, 2.3, 1.5	2.4	13	30
Time	∼30 seconds	∼1 minutes	∼20 minutes	∼2.5 hours

In terms of spatial information, initial experiments suggested that using the centre of the image as a reference point works well for obtaining an initial segmentation in knee MRI. After this, the Euclidean distance transform of the femur and the tibia were used to provide spatial information to refine the segmentation. The weighting for the spatial information, α, is dependent on image resolution and patch size. Table 2 shows the patch sizes and spatial weighting, α, used as well as segmentation examples and computation times on an 8 core 2.7GHz CPU for each resolution. Details of the different resolution levels are shown in table 1. The k parameter for the nearest patches selected was fixed at 15 for all resolutions and 100 atlases were used for resolution levels 4 to 2, whilst 30 were used for level 1. A low value for the k parameter can provide minor speed gains but our experiments suggested $k > 10$ was a minimum requirement for accurate results. Parameters are chosen based on experiments with the training data and are not tuned to the test set.

3.2 Results

The results from the SKI10 grand challenge[1] are presented in table 3, showing promising potential for this framework and demonstrating the possibility of applying patch-based segmentation in images where accurate registration poses a challenge. The framework does not make any prior assumptions about the knees, yet achieves a good score for bone segmentation whilst the score for cartilage segmentation is comparable to the top scoring entries in the challenge.

[1] See `http://www.ski10.org/results.php` for a full list of results from other entries.

Table 3. Overall results from the SKI10 grand challenge showing average surface distance (AvgD), root mean squared surface distance (RMSD), volumetric overlap error (VOE), volumetric difference (VD) and score (Scr) for each resolution level.

	Femur Bone			Tibia bone			Femoral cartilage			Tibial cartilage			
Level	AvgD [mm]	RMSD [mm]	Scr	AvgD [mm]	RMSD [mm]	Scr	VOE [%]	VD [%]	Scr	VOE [%]	VD [%]	Scr	Total Score
4	1.72 ±0.33	2.36 ±0.68	31 ±11	1.74 ±0.68	2.40 ±1.08	8 ±6	75.7 ±3.5	−6.3 ±15.9	38 ±17	76.4 ±9.2	−2.2 ±48.1	27 ±17	25.7 ±7.5
3	1.19 ±0.48	1.91 ±1.00	49 ±15	1.17 ±0.82	1.93 ±1.42	34 ±16	56.2 ±4.3	−1.7 ±15.7	47 ±18	55.9 ±5.7	−0.8 ±24.7	47 ±17	44.3 ±9.8
2	0.96 ±0.50	1.77 ±1.04	56 ±16	0.95 ±0.92	1.84 ±1.59	44 ±18	39.3 ±5.5	0.7 ±13.8	60 ±17	40.5 ±5.5	−1.4 ±17.5	57 ±17	54.3 ±9.3
1	0.92 ±0.52	1.78 ±1.07	57 ±16	0.91 ±0.92	1.87 ±1.60	45 ±20	33.4 ±6.6	0.0 ±13.2	63 ±17	34.3 ±6.4	−1.8 ±16.2	61 ±17	56.5 ±9.2

4 Conclusion

We have presented a new multi-resolution framework for applying patch-based segmentation which is able to segment images without requiring registration, along with an atlas selection method which uses histograms of 3D gradients for image comparison in this context. Additionally, the framework allows a trade-off between segmentation accuracy and speed by selecting the resolution at which the segmentation terminates as well as the number of atlases used. We applied this approach to the MR knee images in the SKI10 grand challenge without any post-processing methods, achieving an overall score of 56.5, with average scores of 62.3 for cartilage and 50.8 for bone. This is the first time a purely patch-based method has been applied to segmenting knee images, producing results that are comparable to many of the other methods used in the challenge despite its simplicity. The scores indicate a promising first application of the proposed framework but potential improvements could be achieved by using more training images as well as as post-processing refinements. Furthermore, this framework could also be coupled with model-based approaches as well as other methods such as EM and graph cut to improve performance.

References

1. Akgül, C.B., Rubin, D.L., Napel, S., Beaulieu, C.F., Greenspan, H., Acar, B.: Content-Based Image Retrieval in Radiology: Current Status and Future Directions. Journal of Digital Imaging 24(2), 208–22 (Apr 2011)
2. 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 (Jul 2009)
3. Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M.: Pyramid Methods in Image Processing (1984)

4. 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)
5. Donoghue, C.R., Rao, A., Pizarro, L., Bull, A.M.J., Rueckert, D.: Fast and accurate global geodesic registrations using knee MRI from the Osteoarthritis Initiative. CVPRW. pp. 50–57 (Jun 2012)
6. Eskildsen, S.F., Coupé, P., Fonov, V., Manjón, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Ø stergaard, L.R., Collins, D.L.: BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59(3), 2362–73 (Feb 2012)
7. Fripp, J., Crozier, S., Warfield, S.K., Ourselin, S.: Automatic Segmentation and Quantitative Analysis of the Articular Cartilages From Magnetic Resonance Images of the Knee. IEEE Transactions on Medical Imaging 29(1), 55–64 (2010)
8. 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), 115–26 (Oct 2006)
9. Heimann, T., Morrison, B.: Segmentation of knee images: A grand challenge. MICCAI Workshop on Medical Image Analysis for the Clinic. pp. 207–214 (2010)
10. Klaeser, A., Marszalek, M., Schmid, C.: A Spatio-Temporal Descriptor Based on 3D-Gradients. Procedings of BMVC pp. 99.1–99.10 (2008)
11. Maurer, C.R., Rensheng, Q., Raghavan, V.: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2), 265–270 (2003)
12. Nyúl, L.G., Udupa, J.K.: On standardizing the MR image intensity scale. Magnetic Resonance in Medicine 42(6), 1072–1081 (1999)
13. 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), 1428–42 (Apr 2004)
14. Rousseau, F., Habas, P., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. IEEE transactions on Medical Imaging 30(10), 1852–1862 (May 2011)
15. Thomson, J.: On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure. Philosophical Magazine 7(39), 237–265 (1904)
16. Torralba, A., Fergus, R., Freeman, W.T.: 80 Million Tiny Images: a Large Data Set for Nonparametric Object and Scene Recognition. IEEE transactions on Pattern Analysis and Machine Intelligence 30(11), 1958–70 (Nov 2008)
17. Tustison, N.J., Avants, B.B., Cook, P.a., Zheng, Y., Egan, A., Yushkevich, P.a., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE transactions on Medical Imaging 29(6), 1310–20 (Jun 2010)
18. Wang, Z., Wolz, R., Tong, T., Rueckert, D.: Spatially Aware Patch-based Segmentation (SAPS): An Alternative Patch-Based Segmentation Framework. Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging, pp. 93–103 (2013)
19. Wolz, R., Chu, C., Misawa, K., Mori, K., Rueckert, D.: Multi-organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases. vol. 7510, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012, pp. 10–17 (2012)