Auxiliary Anatomical Labels for Joint Segmentation and Atlas Registration

Tobias Gass, Gabor Szekely and Orcun Goksel

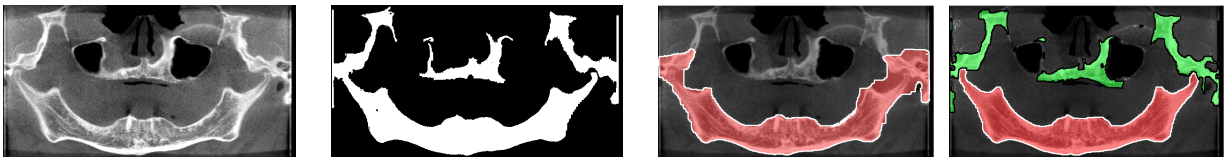
Computer Vision Lab, ETH Zurich, Switzerland.
{gasst, szekely, ogoksel}@vision.ee.ethz.ch

Abstract

This paper studies improving joint segmentation and registration by introducing auxiliary labels for anatomy that has similar appearance to the target anatomy while not being part of that target. Such auxiliary labels help avoid false positive labelling of non-target anatomy by resolving ambiguity. A known registration of a segmented atlas can help identify where a target segmentation should lie. Conversely, segmentations of anatomy in two images can help them be better registered. Joint segmentation and registration is then a method that can leverage information from both registration and segmentation to help one another. It has received increasing attention recently in the literature. Often, merely a single organ of interest is labelled in the atlas. In the presense of other anatomical structures with similar appearance, this leads to ambiguity in intensity based segmentation; for example, when segmenting individual bones in CT images where other bones share the same intensity profile. To alleviate this problem, we introduce automatic generation of additional labels in atlas segmentations, by marking similar-appearance non-target anatomy with an auxiliary label. Information from the auxiliary-labeled atlas segmentation is then incorporated by using a novel coherence potential, which penalizes differences between the deformed atlas segmentation and the target segmentation estimate. We validated this on a joint segmentation-registration approach that iteratively alternates between registering an atlas and segmenting the target image to find a final anatomical segmentation. The results show that automatic auxiliary labelling outperforms the same approach using a single label atlasses, for both mandibular bone segmentation in 3D-CT and corpus callosum segmentation in 2D-MRI.

1 Introduction

In computer vision, in general, and medical imaging, in particular, the segmentation of an image is a common problem that has been studied extensively in the literature. Possible approaches can be categorized by the amount of prior knowledge they use, such as intensity-based segmentation on the one end of the spectrum and model-based segmentation on the other. Intensity-based segmentation typically involves a local classifier to determine the type of tissue that a pixel most likely belongs to. It can additionally regularize such segmentation, for example using Markov random fields (MRF) [1]. This method is typically fast, but lacks anatomical correctness as neither absolute nor relative spatial locations (e.g., shapes) are taken into account. Due to its weak dependency on manual annotations, many recent studies have focused on registration-based segmentation [2] where a single reference (atlas) image is registered to a target image. The resulting transformation is then applied to the labelled atlas, which yields a segmentation of the target image. While being widely used in medical imaging, image registration alone cannot solve the segmentation task by itself as it is known to be an ill-posed problem



(a) target image (b) intensity segmentation (c) single-label (d) + auxiliary label

Figure 1: Illustration of single- vs aux-label joint registration and segmentation. The intensity based segmentation (b) of the target image (a) covers all bone. In a single-label approach (c), the method tries to maximize the overlap between the deformed atlas segmentation of the jaw bone, and bone pixels in the image. The proposed method uses an automatically generated aux-label atlas, which allows to precisely segment the jawbone (c).

[3]. This is because anatomical correspondences, which are not guaranteed to exist, are computed using surrogate criteria such as intensity similarity. In addition, the problem of non-rigid registration is known to be NP-complete [4], thus making approximative algorithms prone to local optima, for example due to poor initialization. Increasing the amount of prior knowledge is a common remedy for such problems, for example by including multiple or statistical atlases [5, 6, 7].

In order to alleviate such problems, joint optimization of segmentation and registration was first proposed by [8]. Such methods allow segmentation and registration processes to mutually benefit from one another. On the one hand, knowing the segmentation of both target and atlas images one can greatly improve their registration. On the other hand, knowing the perfect correspondences between such two images will improve, and even solve completely, the (atlas-based) segmentation problem. Accordingly, several approaches for such joint registration and segmentation have been proposed in the literature, focusing on alternating registration and segmentation (ARS) methods.

An ARS approach that alternates between estimating a rigid deformation using Powells' method and updating the segmentation using iterative conditional modes (ICM) in an MRF was proposed in [8]. Another approach based on MRFs that alternates between solving one MRF to optimize registration parameters and to update segmentation probabilities, and a second MRF to solve the segmentation itself was introduced in [9]. A Bayesian framework to alternate between updating the registration and estimating intensity "nuisance" parameters while marginalizing over possible anatomical labels using an EM algorithm was presented in [10]. There have also been studies on alternating registration and segmentation in variational formulations. Those also rely on iteratively updating the registration and segmentation solutions, for example, by using gradient descent [11, 12].

In many cases, this process can still be ambiguous, as the segmentation criterion may still find some part of the anatomy that matches the given intensity prior, though not being part of the desired target. This might occur for example in the segmentation of single bones, which unavoidably have similar appearance to other bones. This leads to ambiguity in the joint process, leading to complications such as *leakage* as illustrated in Fig. 1(c). In this paper, we propose the use of automatically-generated *auxiliary* anatomical labels to alleviate this problem. Such labels are meant to identify voxels with the appearance of the target anatomy in both the atlas and the target images. We show experimentally that such auxiliary labels improve joint registration and segmentation.

2 Method

Let X be the target and A the atlas image, where both are functions $\Omega \to \mathcal{F}$ from the D-dimensional discretized coordinate domain $\Omega \in \mathbb{N}^D$, to image intensities $\mathcal{F}$. A segmentation $S = \{s_0, \cdots, s_N\}$ is an assignment of one such segmentation label to each pixel of an image. In segmentation, the goal is to find a segmentation S_X of the target image X. Without loss of generality, we assume $s \in \{l_b, l_f\}$, where l_b labels background and l_f labels the target anatomy. In registration, the goal is to find a transformation $\mathcal{T}$ such that the target image X and the transformed atlas image $A(\mathcal{T})$ are similar according to a selected metric. $\mathcal{T}$ can be a *non-rigid* deformation represented by a D-dimensional displacement vector field $T : \Omega \to \mathbb{R}^D$ so that for each pixel/voxel $i \in \Omega$: $\mathcal{T}(i) = i + t_i$, where $\mathcal{T}_i$ and t_i are shorthands for the local deformation $\mathcal{T}(i)$ and displacement $T(i)$. We will also use this notation to denote pixels in an image, e.g. X_i, A_i. Then, $I(\mathcal{T}) = i \to I(i + t_i)\ \forall i$, indicates a deformed image. Values at non-grid locations are linearly interpolated. Both finding a displacement vector field T and a segmentation S_X can be defined as separate energy minimization problems as follows:

$$\hat{S}_X = \arg\min_{S_X} \sum_{i \in \Omega} \left(\psi_i^{\text{seg}}(s_i) + \sum_{j \in \mathcal{N}(i)} \lambda_{\text{seg}}^{ij} \Psi_{ij}^{\text{seg}}(s_i, s_j) \right) \tag{1}$$

$$\hat{T} = \arg\min_{T} \sum_{i \in \Omega} \left(\psi_i^{\text{reg}}(t_i) + \sum_{j \in \mathcal{N}(i)} \lambda_{\text{reg}} \Psi_{ij}^{\text{reg}}(t_i, t_j) \right), \tag{2}$$

where $\mathcal{N}(i)$ denotes the neighbors of a pixel i and ψ, Ψ are unary and pairwise potential functions, and λ are weights. The solutions ensure a smooth labelling of the image and can be found efficiently using graph-based methods such as *α-expansion graphcuts* [13]. We follow standard approaches for both segmentation and registration, using Hounsfield-unit based segmentation potentials as described in [14] for bone segmentation in CT, and learned potential functions in the case of brain MR data. For the registration process, we follow the approach of [2], with normalized cross-correlation as the image similarity metric.

2.1 Joint Registration and Segmentation

As a joint technique, we implement an *alternating registration and segmentation*(ARS) procedure similar to [8]. In ARS, the estimated solution for one subproblem is used as prior knowledge in the other in an iterative manner.

Using MRF approaches for both segmentation and non-rigid registration, we define an additional energy that links the deformed atlas segmentation and the target segmentation, namely *coherence energy* $E^{\text{coh}}(S_X, S_A(\mathcal{T}))$. We compute this energy using a distance weighted overlap penalty between the target segmentation and the deformed atlas segmentation as follows:

$$E^{\text{coh}}(S_A, S_X, T) = \sum_i (1 - \delta(s_i, S_A(\mathcal{T}_i))) \Psi^{\text{coh}}(t_i, s_i) \tag{3}$$

$$\Psi^{\text{coh}}(t_i, s_i) = \frac{\mathcal{D}_{s_i}(\mathcal{T}_i)^2}{2\tau^2}, \tag{4}$$

where δ is the Kronecker-delta, $\mathcal{D}_{s_i}$ denotes the distance transform of S_A with respect to label s_i and τ is a tolerance parameter, which we decrease after each ARS iteration, starting from an empirically-set tolerance of 16mm as can be seen in Fig. 3(right). This decrease in tolerance was motivated by the fact

that the expected difference between deformed atlas segmentation and target segmentation is larger in earlier levels compared to later stages. In experiments, a small tolerance in early levels was observed to undesirably force the target segmentation to be grossly dissimilar from the true segmentation, which in turn leads to a local optimum in the iterative registration process. This was mitigated by the above-mentioned adaptive setting.

The ARS algorithm then alternates between estimating the registration and the segmentation as follows:

$$T^{n+1} = \arg\min_{T} E^{\text{reg}}(X, A, T) + \lambda_{\text{coh}} E^{\text{coh}}(S_A, S_X^n, T)) \tag{5}$$

$$S_X^{n+1} = \arg\max_{S_X} E^{\text{seg}}(X, S_X) + \lambda_{\text{coh}} E^{\text{coh}}(S_A, S_X, T^{n+1}), \tag{6}$$

where λ_{coh} weights the influence of the coherence energy, which is included in the MRF based solver as a unary potential. Note that in the first iteration, no estimate of the segmentation is known and a regular registration is computed. We check for convergence by computing the overlap between S^n and S^{n-1}. The algorithm exits, if this overlap reaches 99% or a maximum of 10 iterations is reached.

2.2 Auxiliary Labels

Naturally, any criterion similar to $E^{\text{coh}}(S_X, S_A(\mathcal{T}))$ will favor a high overlap between a single-label atlas segmentation and any pixels which satisfy an intensity model for this anatomy. This may lead to a sub-optimal solution in the joint registration and segmentation process, where the atlas segmentation is forced to overlap with non-target anatomy pixels. An example is given in Fig. 1, where the mandibular bone is extended into the skull. While this effect may partly be mitigated by strong regularization, this would then not only deteriorate results in regions where high flexibility is necessary, but also could make the algorithm more sensitive to initialization.

We thus propose *auxiliary anatomical labels* in both the atlas and the target segmentation in order to prevent such false positives in ARS segmentation. Such auxiliary labels can be obtained automatically using the following procedure:

1. An automatic binary segmentation $\bar{S}_A$ of the atlas image is computed using Eq. (1).
2. The ground-truth atlas segmentation is dilated and subsequently subtracted from the estimate $\bar{S}_A$. Such dilatation helps remove minor artefacts near the boundary of the ground-truth segmentation.
3. The auxiliary label l_a is then assigned to the remaining voxels in $\bar{S}_A$ before combining it with the ground-truth S_A to generate an aux-label atlas segmentation.

An example of this process is shown in Fig. 2.

2.2.1 Coherence Potentials foe Auxiliary Labels

While the anatomy represented by the auxiliary label is expected to provide valuable information to an ARS algorithm, it is less reliable in comparison to a manual annotation for two reasons: (i) The auxiliary labelling is obtained automatically and thus prone to errors due to imaging artefacts. (ii) While the primary label was chosen carefully by a medical expert, our automatically labelled structures may not represent anatomies for which correspondences exist across different patients. Therefore, such auxiliary labels are expected inherently to have less *coherence* between the deformed atlas segmentation

(a) atlas image (b) automatic segmentation (c) manual segmentation (d) aux-label segmentation

Figure 2: Pipeline for obtaining the auxiliary-labelled atlas segmentation. The original image (a) is segmented using the segmentation sub-graph yielding a binary segmentation (b). The known ground-truth segmentation (c) is then removed from an eroded version of (b) in order to obtain the second label segmentation, which is subsequently diluted. In a last step,the ground-truth and the second label segmentation are combined, yielding the final aux-label segmentation (d).

and the target segmentation estimate. We accommodate this by augmenting the coherence potential as follows:

$$\Psi^{\text{coh}}_{\text{aux}}(t_i, s_i) = \begin{cases} \Psi^{\text{coh}}(t_i, s_i) & \text{if } (s_i \equiv l_f) \vee (S_A(\mathcal{T})_i \equiv l_f), \\ \min(1, \Psi^{\text{coh}}(t_i, s_i)) & \text{otherwise.} \end{cases} . \tag{7}$$

This function essentially truncates the coherence potential to a maximum of 1, if neither the target segmentation label nor the deformed atlas segmentation label is the foreground label l_f. This limits the cost of labelling pixels in the target image as background, when the deformed atlas segmentation is labelled as l_a. Analogously, the cost of labelling pixels as auxiliary anatomy, when the deformed atlas segmentation is labelled as background, is then also limited.

3 Experimental Evaluation

We evaluated our method on two datasets: a set of 70 mid-saggital slices of 2D brain MR, where the task is to segment the corpora callosa, and a set of 14 3D CTs of the head, where the task is to segment the mandibular bone. The mandibular bone dataset represents a typical clinical problem, and the corpus callosum dataset provides an opportunity to evaluate our algorithm on a significantly larger dataset and a different modality. Both datasets were rigidly registered to a common reference frame, and gold-standard segmentations have been provided by experts. We report Dice's coefficient (DICE), Hausdorff distance (HD) and mean surface distance (MSD) as quantitative measures of segmentation accuracy. We optimized the parameters of each method separately using a small subset of the images. Additionally, we report results for pure MRF segmentation (MRF-Seg.), and for pure MRF registration which is our implementation of [2] (MRF-Reg.).

Mandibular Bone Segmentation. A dataset of 15 3D CT scans of the head with 160x160x120 px^3 dimensions and 1 mm^3 voxel size was used for our validation, with the mandibles in each segmented by an expert. Due to older age group, the amount of teeth varies largely in this set, which in addition to artefacts from metal implants makes both automatic registration and segmentation of the mandibles a challenging task. We performed a leave-one-out (l1o) experiment by using each of the 15 images as the atlas image in turn to segment the remaining 14 samples, resulting in 210 experiments for each method. In Fig. 3(left), we present the average performance values over all experiments for the above-mentioned methods. Note that the pure MRF-based bone segmentation method (MRF-Seg.) is the method presented in [14], for which no atlas image is needed and thus only 15 experiments were required for

	2D MRI			3D CT		
method	DICE	MSD	HD	DICE	MSD	HD
MRF-Seg.	0.837	0.500	8.34	0.90	3.39	17.00
MRF-Reg.	0.918	0.093	2.85	0.849	0.32	11.69
ARS (single-label)	0.918	0.084	2.31	0.924	0.16	9.1
ARS (aux-labelled)	0.925	0.064	2.11	0.933	0.13	7.88

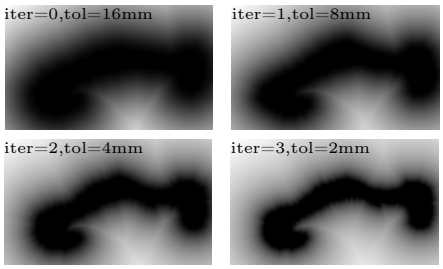


Figure 3: Left: Average segmentation quality measures for the methods on both datasets. Right: Progression of the coherence potential for a given deformation and decreasing tolerance.

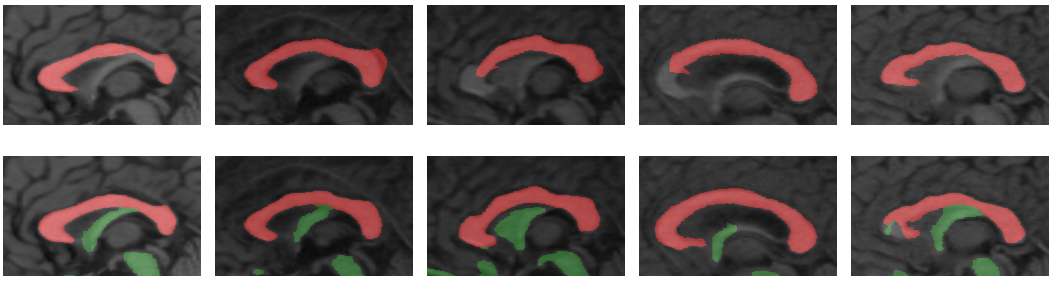


Figure 4: Example segmentations of the corpus callosum in 2D MR. Top row shows single-label ARS segmentation estimate, bottom row with the auxiliary label. Primary label is shown in red, auxiliary label is colored green. In most cases, the auxiliary-labelled estimate yields a better segmentation in comparison to single-label ARS. The right-most column shows a failure case, where the auxiliary segmentation label is mis-matched, causing an incorrect alignment.

this method. Also, [14] merely finds a separate label for each connected bone region, assuming the target anatomy to be the largest connected region. However, this is not true for the mandible in the head CT scans. Therefore, we had to select the label for the mandible in a manual post-processing step, positively biasing the results in favor of MRF-Seg [14].

Corpus Callosum Segmentation. A second experiment was performed using 70 mid-saggital slices of 2D MR images of the brain with $0.33\,\mathrm{mm}^2$ pixel resolution. After rigid pre-alignment, each image was cropped to a fixed $200\text{x}120\,\mathrm{px}^2$ region of interest. The total number of leave-one-out experiments for each method is then 4830, where we performed the same experiments as above, with the exception of the segmentation potential functions having been replaced by intensity based classifiers. We used Gaussian mixture models for foreground/background with four densities for the segmentation unary potential and random forests [15] for the pairwise potentials, both trained using the atlas image. The negative logarithm of the classification probabilities were used as unary and pair-wise energies in the segmentation graph. We report the average metrics similarly in Fig. 3(left), and show sample results in Fig. 4.

3.1 Discussion

For both the corpus callosum and the mandibular bone segmentation tasks, aux-label ARS outperforms single-label ARS significantly ($p \leq 1.7e^{-4}$). This gain comes at some moderate additional cost, as aux-label ARS requires the additional computation of the auxiliary label distance transform in the coherence

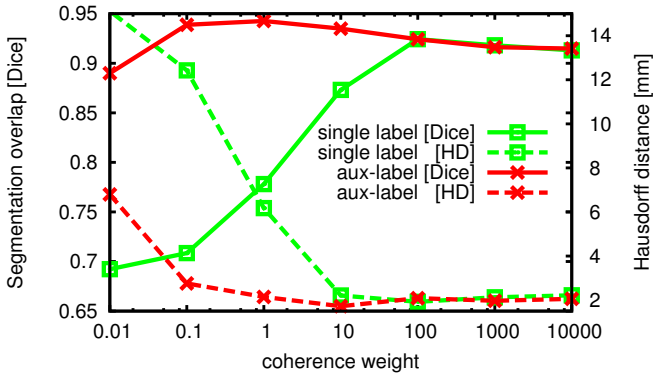


Figure 5: Effect of coherence potential weight on single- and aux-label ARS.

potential for both the segmentation and the registration steps, and also an additional segmentation label needs to be generated in the segmentation step. We solve the aux-label segmentation problem using the efficient alpha-expansion graph-cuts algorithm [13], which can guarantee that the solution is within a known factor of the global optimum. The following observations are made when comparing single- and aux-label ARS with respect to changing the free parameter λ_{coh}. Firstly, the optimal λ_{coh} is an order of magnitude larger for the single-label ARS compared to its auxiliary-labelled version. Additionally, the optimal segmentation weights are slightly lower for the single-label ARS, causing it to more strongly rely on the registration information. This is indeed expected, as the segmentation sub-problem is essentially ill-posed when using only one label. In contrast, aux-label ARS is much less susceptible to sub-optimal parametrization in comparison to single-label ARS as can be seen from the narrower range in both Dice's coefficient and Hausdorff distance. This demonstrates that the aux-label ARS formulation is indeed better suited for joint segmentation and registration problems as in this paper.

4 Conclusions

In this work, we have presented a novel approach using automatically generated auxiliary anatomical labels for joint segmentation and atlas registration. A distance-weighted overlap energy was proposed to couple the segmentation and registration process in an iterative fashion. We show experimentally that automatic auxiliary labelling helps improve segmentation quality compared to joint segmentation and registration using single target label alone. Although the presented method focuses on segmentation as a pixel-labelling problem, we expect similar behaviour for other segmentation methods such as active contours. Here, evolving a target contour simultaneously to auxiliary contours may help to avoid poor local optima which can result from initialisation problems or image artefacts.

References

[1] Boykov, Y., Funke-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation. Int J Comput Vision **70**(2) (2006) 109–131

[2] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med Image Anal **12**(6) (2008) 731–741

[3] Fischer, B., Modersitzki, J.: Ill-posed medicine - an introduction to image registration. Inverse Problems **24**(3) (2008) 1–19

[4] Keysers, D., Unger, W.: Elastic Image Matching is NP-complete. Pattern Recognition Letters **24** (2003) 445–453

[5] Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? Handbook of Biomedical Image Analysis (2005) 435–486

[6] 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) (2006) 115–26

[7] Glocker, B., Komodakis, N., Paragios, N., Glaser, C., Tziritas, G., Navab, N.: Primal/dual linear programming and statistical atlases for cartilage segmentation. In: MICCAI. Volume 10. (2007) 536–543

[8] Wyatt, P., Noble, J.: MAP MRF joint segmentation and registration of medical images. Med Image Anal **7**(4) (2003) 539–552

[9] Xiaohua, C., Brady, M., Rueckert, D.: Simultaneous segmentation and registration for medical image. In: MICCAI. (2004) 663–670

[10] Pohl, K., Fisher, J., Levitt, J., Shenton, M.: A unifying approach to registration, segmentation, and intensity correction. In: MICCAI. (2005) 310–318

[11] Ghosh, P., Sargin, E., Manjunath, B.: Generalized simultaneous registration and segmentation. In: CVPR. (2010) 1363–1370

[12] Schmidt-Richberg, A., Ehrhardt, J.: Lung Registration with Improved Fissure Alignment by Integration of Pulmonary Lobe Segmentation. MICCAI **7511** (2012) 74–81

[13] Boykov, Y., Veksler, O., Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans Pattern Anal Mach Intell **23**(11) (2001) 1222–1239

[14] Furnstahl, P., Fuchs, T., Schweizer, A., Nagy, L., Székely, G., Harders, M.: Automatic and robust forearm segmentation using graph cuts. In: ISBI. (2008) 77–80

[15] Breiman, L.: Random forests. Machine learning **45**(1) (2001) 5–32