

Automated Skeleton Based Multi-modal Deformable Registration of Head&Neck Datasets

Sebastian Steger and Stefan Wesarg

Cognitive Computing & Medical Imaging, Fraunhofer IGD, Darmstadt, Germany
{sebastian.steger,stefan.wesarg}@igd.fraunhofer.de

Abstract. This paper presents a novel skeleton based method for the registration of head&neck datasets. Unlike existing approaches it is fully automated, spatial relation of the bones is considered during their registration and only one of the images must be a CT scan. An articulated atlas is used to jointly obtain a segmentation of the skull, the mandible and the vertebrae C1-Th2 from the CT image. These bones are then successively rigidly registered with the moving image, beginning at the skull, resulting in a rigid transformation for each of the bones. Linear combinations of those transformations describe the deformation in the soft tissue. The weights for the transformations are given by the solution of the Laplace equation. Optionally, the skin surface can be incorporated. The approach is evaluated on 20 CT/MRI pairs of head&neck datasets acquired in clinical routine. Visual inspection shows that the segmentation of the bones was successful in all cases and their successive alignment was successful in 19 cases. Based on manual segmentations of lymph nodes in both modalities, the registration accuracy in the soft tissue was assessed. The mean target registration error of the lymph node centroids was 5.33 ± 2.44 mm when the registration was solely based on the deformation of the skeleton and 5.00 ± 2.38 mm when the skin surface was additionally considered. The method's capture range is sufficient to cope with strongly deformed images and it can be modified to support other parts of the body. The overall registration process typically takes less than 2 minutes.

Keywords: Image Registration, Head&Neck, Multi-Modal, Multi-Rigid.

1 Introduction

The registration of intra subject head&neck 3D datasets is required in many clinical applications. In image guided radiation therapy it enables adapting a previously generated plan to the patient's pose during intervention. Furthermore, it can be used for image fusion of different imaging modalities like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Thus, the clinical target volume and organs at risk can be delineated in the more suitable modality and propagated to the other modality, thereby improving the treatment plan. If spatial correspondence by the means of image registration is available, such a plan can also be propagated to a follow up CT scan which facilitates the plan

adaption. Beyond that, computer aided diagnosis systems can benefit from such an image fusion as they can incorporate different modalities for image feature extraction from clinically relevant targets such as tumors or lymph nodes.

Whereas rigid registration works well for the head, it is not able to cope with the neck due to possibly different positions of the spine. Intensity based deformable registration techniques typically have a smaller capture range and may produce poor results in regions with low contrast/signal or heavy imaging artifacts.

A popular approach to overcome these limitations is to exploit that the skeleton is the supporting structure of the soft tissue. The bones are treated as individual rigid bodies and deformations are allowed in the soft tissue only. Originally proposed by Little et al. [9], many other works are based on this principle. It is mostly applied to the registration of the spine [7][6][15][14], but also to the head&neck [5][1], the brain[3], the hand [10] and even to the whole body [8]. On Micro-CTs it is used for the registration of the lower limbs [11] and the whole body [4][13] of mice.

One of the key challenges of this principle is the identification of the individual rigid parts, i.e. the bones. Whereas most methods [1][3][5][9][11][15] depend on manual or interactive segmentation, [6][7] pursue semi-automatic approaches while [10][14] present very specific automatic heuristics. In [4], an articulated atlas is used. Methods not requiring explicit segmentation [8][13] of the individual bones can only be used if both images are CT scans from which the bone surface can easily be extracted.

During the registration of the individual bones the articulation is only considered in a few cases [4][5][10][11]. The deformation of the surrounding soft tissue can be computed solely based on the rigid transformations using thin plate splines (TPS) [7], linear combinations of transformations [9][14][10], a Log-Euclidean framework [3] or finite element methods (FEM) [5][1]. Alternatively, the rigid parts can be incorporated as constraints [15] or serve as an initialization [13][8] for an overall deformable intensity based registration. Rarely, the evaluation of the registration accuracy is carried out quantitatively. In [1], the center of volume error and dice similarities are reported for manually delineated targets whereas the distance of 20 landmarks is reported in [5]. Both evaluations are based on only 4 pairs of images.

In this paper, we present a fully automated registration method based on multi-rigid registration of the skeleton for 3D images of the head&neck. One of the images must be a CT image – the *fixed image* – but the other one – the *moving image* – can be from a different modality for which rigid registration methods exist or an image of the same subject acquired at a different point in time. After an overall mutual information based rigid registration of the head, the articulated atlas presented in [12] is used to obtain a segmentation of the individual bones from the CT image. These bones are then successively rigidly registered with the moving image resulting in a rigid transformation for each of them. Linear combinations of those transformations describe the soft tissue deformation. Optionally, the skin surface can be incorporated. The approach is quantitatively evaluated on 20 CT/MRI pairs of clinically acquired head&neck datasets.

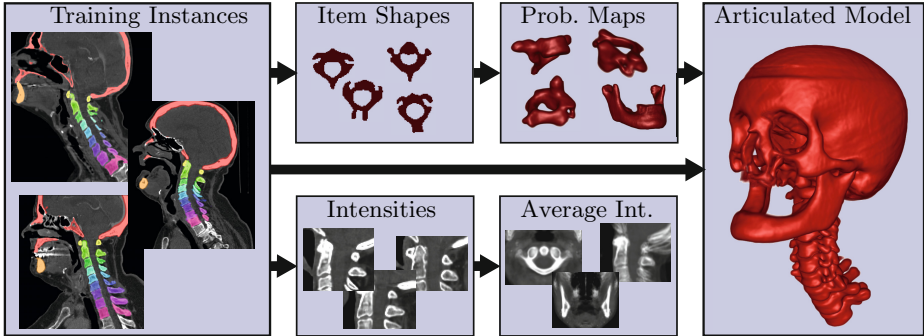


Fig. 1. Training of the articulated atlas for segmentation of bones from the fixed image

2 Methods

Articulated Atlas for CT Bone Segmentation. Due to its high contrast, bone can easily be segmented in CT images. However, since different bones have a similar appearance and are close to each other (e.g. the vertebrae), simple image segmentation techniques like thresholding are not able to distinguish neighboring bones from each other. Therefore, we use the articulated atlas presented in [12]. It is capable of jointly segmenting the skull, mandible, the cervical vertebrae and the two upper thoracic vertebrae. The atlas was created from head&neck CT datasets of 15 patients for which all bones had been labeled manually. Assuming bilateral symmetry of the bones, their right/left flipped versions were also used, resulting in a total of 30 training instances.

Unlike Baiker et al.’s articulated atlas [4], the articulation of the rigid parts is not explicitly modeled but learned from the training images. Therefore, particular attention was paid to cover a wide range of possible poses of the spine when selecting the training images.

For each type of bone, the probability map indicating the membership to the item along with the average intensities was computed in a coordinate system normalized by translation, rotation and scaling. Each bone item has those 7 degrees of freedom. Based on Alexa’s *Linear Combination of Transformations* [2] a 7-dimensional vector is used to uniquely describe the pose of an item with respect to the model’s global translation and rotation as a linear combination of 7 basis transformations. For a total of m bone items, the overall pose of each training instance j is described by a $7m$ -dimensional vector \mathbf{x}_j .

The space of all possible articulations is then described by applying Principal Component Analysis (PCA) on the training articulation $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$. An arbitrary articulation \mathbf{x} can be described with a $k \ll 7m$ dimensional vector \mathbf{b} as $\mathbf{x} = \bar{\mathbf{x}} + A \cdot \mathbf{b} + \mathbf{r}$, where $\bar{\mathbf{x}}$ is the average articulation of all training instances \mathbf{x}_j and A is a matrix consisting of the eigenvectors of the covariance matrix. The more likely an articulation \mathbf{x} is, the smaller the magnitude of \mathbf{b} and \mathbf{r} will be.

The atlas is adapted to a CT image by minimizing an energy functional depending on the 6 dimensional vector \mathbf{t} describing a global rigid transformation and the $7m$ dimensional articulation parameter vector \mathbf{x} :

$$E(\mathbf{x}, \mathbf{t}) = E_{\text{external}}(\mathbf{x}, \mathbf{t}) + \lambda E_{\text{internal}}(\mathbf{x}). \quad (1)$$

The external energy E_{external} ensures gray value similarity between the test image and the trained intensities of the atlas items and that the atlas items are close to high CT intensities (i.e. bones). The internal energy E_{internal} ensures that the atlas is within or at least close to the trained articulation space. The empirically determined parameter λ balances the external and internal energy. A gradient descent optimizer first finds the global transformation parameters \mathbf{t} and then the articulation and external parameters \mathbf{x}, \mathbf{t} jointly. Once the articulated model converged, the segmentations S_j for each bone item are extracted based on the probability maps and the CT intensities. Please refer to [12] for details.

Successive Rigid Alignment of the Bones. Based on the segmentations S_j of the bones in the CT image, they are rigidly aligned to the moving image by maximizing mutual information with a gradient descent optimizer. In order to incorporate not only the rather homogeneous part inside the bone, but also the texture at its boundary, the fixed image samples are taken from a slightly (5 mm) dilated version of S_j during the metric computation process as suggested in [14]. The success of local optimization techniques finding the desired local maximum depends on an initialization close to that very maximum. For highly deformed head&neck images, the same vertebra can be far away in both images and thus an independent registration is not feasible. Therefore, we exploit the connectivity of the bones expressed in the order in which the items are aligned. We start with the skull and use its resulting rigid transformation to initialize the registration of the mandible and the topmost vertebra (C1). The result of the latter is then used to initialize the next vertebra (C2) and so on, resulting in a top-down strategy. At the end, m rigid transformations T_j are obtained, one for each bone item j . Effectively denoting an overall transformation as a concatenation of several transformations enlarges the capture range.

Soft Tissue Motion Coupling. Two main constraints are applied when computing the transformation field $T(\mathbf{x})$ for each point in the fixed image: i) The known transformations in the rigid parts shall be incorporated (i.e. $T(\mathbf{x}) = T_j \Leftrightarrow \mathbf{x} \in S_j$) and ii) the transformation field shall be continuous (i.e. $T(\mathbf{x} + \epsilon) \approx T(\mathbf{x})$ for small $|\epsilon|$). Note that not only the translational part but also the rotational part shall be continuous, which is advantageous as argued in [9].

We denote the transformation at an arbitrary location \mathbf{x} as a linear combination of the known transformations T_j , again using the approach presented in [2]: $T(\mathbf{x}) = \sum_{j=1}^m \Phi_j(\mathbf{x})T_j$. To achieve continuity, each component of the m -dimensional coefficient field shall satisfy Laplace's partial differential equation (PDE): $\Delta\Phi_j = 0$, where Δ is the Laplace operator. Dirichlet boundary conditions $\Phi_j(\mathbf{x}) = \{1 \Leftrightarrow \mathbf{x} \in \partial S_j; 0 \Leftrightarrow \mathbf{x} \in \partial S_{i \neq j}\}$ are used at the boundaries of the bone items ∂S_j enforcing our first constraint. In a first step, the

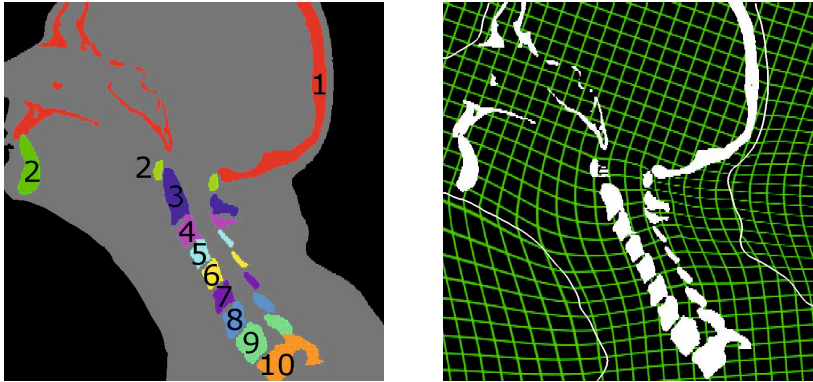


Fig. 2. Domain of the PDE in gray, registration order (left), deformed pattern (right)

domain Ω of the PDE is restricted to the body (gray region in Fig. 2) which is obtained by global thresholding. This ensures that transformations are only propagated within the body and undesired influences – such as from the chin directly into the thorax – are prevented. Here, Neumann boundary conditions are deployed ($\frac{\partial \Phi_j(\mathbf{x})}{\partial x_i} \Big|_{\mathbf{x} \in \partial \Omega} = 0$). The solution of the PDE is found by solving the linear equation system resulting from finite differences approximation. In a second step, the domain is extended to the rest of the image (black region in Fig. 2) and previously obtained transformations are used as boundary condition.

Further Refinement. Since the computation of the soft tissue deformation solely depends on the rigid transformations of the bones, the continuity assumption may not result in the desired accuracy, especially if large deformations are present. One can imagine many different solutions to overcome this issue, such as a potentially more realistic deformation model (e.g. finite element methods [5][1]) or the incorporation of soft tissue image intensities into the registration process with rigidity constraints for the bones [15]. However, the approach we pursue is including further landmarks – the body surface – as additional constraints. Extracting the body surface in both images is done by simple thresholding. Since only small differences between both surfaces are expected, correspondences are found along the normal vectors of the fixed image’s surface which are then used to correct the translational part of the deformation field in the entire soft tissue.

3 Experiments and Results

The presented fully automated registration method was evaluated on 20 intra-subject pairs of CT/MRI head&neck images of oral cancer patients acquired in clinical routine. The spacing between CT slices was 1 – 2 mm. The MRI images consist of T1 weighted, fat saturated sagittal slices with a slice gap of 3 – 5 mm. Some of the images showed heavy imaging artifacts.

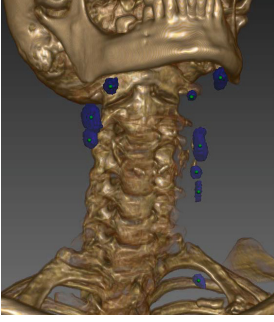


Fig. 3. Evaluation targets

Table 1. The minimum, maximum and average target registration error of lymph node centroids in mm

TRE	Rigid	Bone	Bone+Skin
Avg.	7.69 ± 4.05	5.33 ± 2.44	5.00 ± 2.38
Min.	1.33	0.37	0.92
Max.	61.21	26.20	25.32

At first, the segmentation of the bones in the CT images was assessed. Since the segmentation accuracy is not critical for the subsequent steps, visual inspection is sufficient to assess if all articulated atlas items ended up at the correct target structures. This was the case for all 20 CT-Images, indicating the robustness of the deployed articulated atlas. The successive rigid alignment was assessed in the same way. Except for one single MRI image, all rigid structures were aligned correctly. The reason for the failure in that one case was a very low MRI signal located directly in the spine. Thus, one vertebra converged to the wrong local minimum, resulting in a design based misalignment of all subsequently aligned vertebrae.

Then, in a second experiment, the soft tissue accuracy was assessed for the 19 successful aligned image pairs. Since the evaluation was based on clinical images, neither a ground truth deformation field nor artificial corresponding landmarks were available. Therefore, we chose to measure the target registration error (TRE). As target landmarks we used the centroid of lymph nodes, which were manually delineated in both imaging modalities. On the average we found 5.2 clearly corresponding targets per dataset. The reason we chose lymph nodes is that they are clinically relevant, reasonably well distributed in the soft tissue of the head&neck region (see Fig. 3), locally bound and visible in both imaging modalities. Due to the images' different resolutions and the resulting segmentation differences the centroid appeared to be more robust than e.g. the average surface distance. The average and maximum TRE was separately computed for the bone-aligned and the surface-refined images as well as for the rigidly head-aligned image for comparison. The results are presented in Table 1 and an example for a registration result for a strongly deformed pair of images can be seen in Fig. 4. Whereas targets close to the bones resulted in high accuracies by design, the worst accuracy were achieved very far away from the head. There, different positions of the shoulder – not part of the skeleton model – had an impact, explaining the high maximum TREs. Further inaccuracies can be attributed to inaccurate segmentations due to the low inter slice resolution of the MRI images. Since only primitive image processing techniques were deployed to

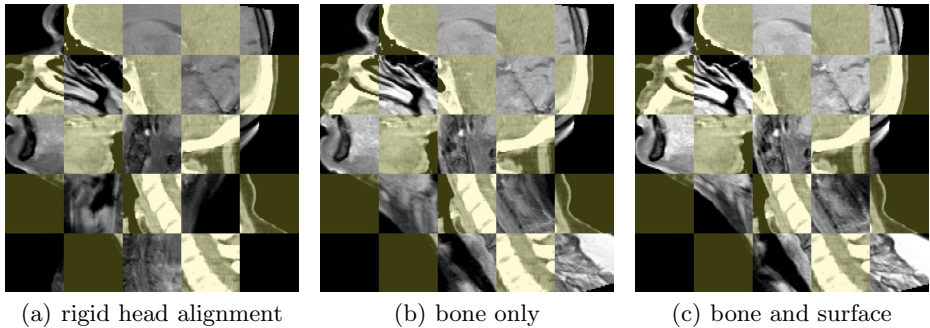


Fig. 4. CT/MRI Registration of a strongly deformed neck (fixed CT colored)

detect the skin surface, this step was locally not successful for some MRI images, resulting in geometric distortion. Model based approaches may be required.

To give a rough indication about the runtime performance of our method, we measured the duration of our prototypical implementation for each individual step on a Intel Core i7-2600K CPU, 3.4Ghz machine with 16GB memory. On the average it took 107s to finish the overall registration, where the rigid registration accounted for 14%, the bone segmentation for 32%, the successive alignment for 16%, the soft tissue motion coupling for 22% and the refinement step for 16%. We are convinced that a considerable speedup is achievable with an optimized implementation exploiting parallelization.

4 Conclusion

In this paper we presented a skeleton based deformable registration approach and applied it to head&neck CT/MRI datasets. Unlike existing approaches, it is fully automated, supports all modalities for which rigid registration techniques exist, is able to cope with strong deformations and was evaluated quantitatively on images acquired in clinical routine. The mean target registration error of the lymph node centroids was 5.33 ± 2.44 mm when the registration was solely based on the deformation of the skeleton and 5.00 ± 2.38 mm when the skin surface was additionally considered. Whereas the articulated atlas enables the automation, the support of strongly deformed image pairs is ensured by the concatenation of several transformations to a single transformation of greater magnitude during the successive alignment of the individual bone structures. With a runtime of less than 2 minutes it is suitable for the use in clinical routine.

Rebuilding the articulated atlas, the approach can be extended to other parts of the body (e.g. the thorax or the pelvis). Without the refinement step, the presented approach has another interesting property when applied to follow up CT scans. Under the assumption that the bones remain constant over time, the method does to some extent compensate for deformations induced by a change in

position only, but not for changes in anatomy (e.g. tumor growth or surgery) because the soft tissue image intensities are not considered. Whereas rigid registration has this property as well, almost no deformable intensity based registration approaches are able to distinguish between those two sources of deformation at all. This may be disadvantageous in some applications (e.g. propagating a radiation therapy plan), but it enables detecting differences in follow up images.

Future work includes increasing the robustness for images with imaging artifacts, deploying more realistic soft tissue motion coupling, applying the presented scheme to other body parts and the automated detection of differences.

References

1. Al-Mayah, A., et al.: Biomechanical-based image registration for head and neck radiation treatment. *Phys. Med. Biol.* 55(21), 6491 (2010)
2. Alexa, M.: Linear combination of transformations. In: *SIGGRAPH 2002*, pp. 380–387. ACM, New York (2002)
3. Arsigny, V., et al.: A fast and log-euclidean polyaffine framework for locally linear registration. *J. Math Imaging Vis.* 33, 222–238 (2009)
4. Baiker, M., et al.: Fully automated whole-body registration in mice using an articulated skeleton atlas. In: *Proc IEEE Int. Symp. Biomed. Imaging*, pp. 728–731 (2007)
5. du Bois dAische, A., et al.: Estimation of the deformations induced by articulated bodies: Registration of the spinal column. *Biomed Signal Proces.* 2(1), 16–24 (2007)
6. Hu, Y., Haynor, D.R.: Multirigid registration of mr and ct images of the cervical spine. In: Fitzpatrick, J.M., Sonka, M. (eds.) *Medical Imaging 2004: Image Processing*, vol. 5370, pp. 1527–1538. SPIE (2004)
7. Huesman, R., et al.: Deformable registration of multi-modal data including rigid structures. In: *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 3, pp. 1879–1882 (2002)
8. Li, X., Peterson, T.E., Gore, J.C., Dawant, B.M.: Automatic Inter-subject Registration of Whole Body Images. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) *WBIR 2006*. LNCS, vol. 4057, pp. 18–25. Springer, Heidelberg (2006)
9. Little, J., et al.: Deformations incorporating rigid structures. In: *Proc. Math. Methods in Biomed. Image Anal.*, pp. 104–113 (June 1996)
10. Martín-Fernández, M.A., et al.: Automatic articulated registration of hand radiographs. *Image Vision Comput.* 27(8), 1207–1222 (2009)
11. Papademetris, X., Dione, D.P., Dobrucki, L.W., Staib, L.H., Sinusas, A.J.: Articulated Rigid Registration for Serial Lower-Limb Mouse Imaging. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 919–926. Springer, Heidelberg (2005)
12. Steger, S., et al.: Articulated atlas for segmentation of the skeleton from head & neck ct datasets. In: *Proc IEEE Int. Symp. Biomed. Imaging*, pp. 1256–1259 (2012)
13. Suh, J.W., et al.: A non-rigid registration method for serial lower extremity hybrid spect/ct imaging. *Med. Image Anal.* 15(1), 96–111 (2011)
14. Čech, P., et al.: Piecewise rigid multimodal spine registration. In: Handels, H., Ehrhardt, J., Horsch, A., Meinzer, H.-P., Tolxdorff, T. (eds.) *Bildverarbeitung für die Medizin 2006*, pp. 211–215. Informatik aktuell, Springer, Heidelberg (2006)
15. Wang, K., He, Y., Qin, H.: Incorporating Rigid Structures in Non-rigid Registration Using Triangular B-Splines. In: Paragios, N., Faugeras, O., Chan, T., Schnörr, C. (eds.) *VLSM 2005*. LNCS, vol. 3752, pp. 235–246. Springer, Heidelberg (2005)