Tensor Controlled Local Structure Enhancement of CT Images for Bone Segmentation

C-F. Westin, S. Warfield, A. Bhalerao, L. Mui, J. Richolt, and R. Kikinis

Surgical Planning Lab, Radiology, Brigham and Women's Hospital, westin@bwh.harvard.edu, Harvard Medical School, Boston

Abstract. This paper addresses the problem of segmenting bone from Computed Tomography (CT) data. In clinical practice, identification of bone is done by thresholding, a method which is simple and fast. Unfortunately, thresholding alone has significant limitations. In particular, segmentation of thin bone structures and of joint spaces is problematic. This problem is particularly severe for thin bones such as in the skull (the paranasal sinus and around the orbit). Another area where current techniques often fail is automatic, reliable and robust identification of individual bones, which requires precise separation of the joint spaces. This paper presents a novel solution to these problems based on three-dimensional filtering techniques. Improvement of the segmentation results in more accurate 3D models for the purpose of surgical planning and intraoperative navigation.

1 Introduction

Three-dimensional renderings of bone that are based on thresholding are currently available on most state-of-the-art CT consoles. Unfortunately, using only thresholding to segment the CT data leads to suboptimal and unsatisfying results in the vast majority of the cases. It is seldom possible to automatically separate the different bones adjacent to a given joint (e.g. femur and pelvis) or different fracture fragments. Problems also exist in areas with very thin bone, such as the paranasal sinuses and around the orbits. So far bone and joint disorders are primarily visualized by plain radiographs or cross sectional CT and MR images. Neither of them are intuitive tools for displaying the patient's anatomy, especially if the spatial relationships of different structures are more complex. 3D visualization can be very helpful for example in cases of pelvic, hip and shoulder disorders. Also pathoanatomic findings in areas with small joints (such as the wrist, hindfoot and spine) are much easier to understand when displayed in 3D.

The adaptive filtering algorithm presented in [1] has been implemented for the specific needs of CT segmentation. The major change presented here is adding the ability to create filters on grids which have different spatial resolutions in different directions to cope with the large out-of-plane/in-plane voxel ratio. Additionally, filters that have a well tuned frequency characteristic in all directions of the data is important as many of the structures of interest have a size close to the signal spacing.

In this paper an adaptive filtering algorithm is designed to emphasize structures of interest in CT data to make segmentation of bone easier. This is an extension of our work presented in [2] where information of locally planar structures were used in an adaptive thresholding algorithm.

Most image enhancement algorithms presented in the literature have design criteria based on characteristics of the perceived quality, such as that the human eye is more sensitive to noise in dark regions and less sensitive to noise in bright and structured ones. Since our goal is not to improve the human visual perception of the data but to perform filtering to improve segmentation, our constraints are slightly different. This paper is based on an adaptive filtering algorithm proposed by Knutsson, Wilson and Granlund [3, 4, 1]. The reason for choosing this particular image enhancement algorithm is that extra edge sharpening can be obtained while smoothing along structures in a very controlled way. Algorithms that are based on anisotropic diffusion information [5,6] or level set theory (so called "geometric diffusion") [7,8] also perform smoothing along structures which to some extent sharpen edges. However, how best to incorporate a stable and controllable sharpening into schemes based on anisotropic diffusion is an open question. Some sharpening can be obtained by running the diffusion equation "backwards", but causes instability and may give unpredictable results. In the literature several methods have been reported to obtain more robust and accurate bone segmentation based on for example region growing and deformable contours. In principle, any of these methods can be used in conjunction with the method proposed in this paper.

2 Algorithmic description

The approach we have taken is to filter the CT data using adaptive filters for the purpose of enhancing the local image structure. The organization of this section is that we begin with presenting a list of the main steps involved in the segmentation scheme. A detailed description of each of the items in the list then follows.

- 1. Design filters for the data grid which is not equally spaced in all directions.
- 2. Resample the data in the slice direction to obtain dense enough sampling to describe a classification result as a binary label map.
- 3. Estimate local image structure and represent it by tensors, and use the local structure tensors to locally control the adaptation of the filters.
- 4. Segment the adaptively filtered result.

Step 1: Filter design on non-cubic grids

In CT data, the voxels are in general non-cubic, i.e. the distance between the samples are not the same in all direction. The in-plane resolution in CT image volumes is often more than a factor of 5-10 higher than the through-plane resolution, e.g. $0.5 \times 0.5 \times 3 \text{ mm}^3$. Data sampled like this has a frequency content

that varies with the direction of the data. In order to get invariance of a filter response, i.e. to get the same sensitivity of structures independent of the orientation of the structure, the frequency characteristic of the filters used must be altered to compensate for the differences in the data. Spatially designed kernels can be compressed in the slice direction before being sampled. However, when designing filters in the frequency domain there is, in most cases, no closed spatial form of the filter available. This is particularly true when other additional constraints, such as locality of the filters, are introduced during the optimization procedure when creating the filters. This means that there is not a spatial filter function that can be resampled, and any compensation of the non-isotropic sampling grid must first be applied in the frequency domain. The relation of a filter, $f(\mathbf{x})$, and linearly transformed version, $g(\mathbf{x}) = f(\mathbf{A}\mathbf{x})$, in the Fourier domain is

$$G(\mathbf{u}) = \frac{1}{|\det \mathbf{A}|} F(|\mathbf{A}^T|^{-1} \mathbf{u})$$
(1)

A scaling of the spatial grid along the coordinate axes can be written,

$$\mathbf{A} = \begin{pmatrix} a \ 0 \ 0 \\ 0 \ b \ 0 \\ 0 \ 0 \ c \end{pmatrix} \tag{2}$$

where a, b and c are the scaling factors. According to Equation 1, this implies the following relation of the two Fourier transforms F and G.

$$G(u_1, u_2, u_3) = \frac{1}{abc} F(\frac{u_1}{a}, \frac{u_2}{b}, \frac{u_3}{c})$$
(3)

Step 2: Resampling

The output from segmenting gray-level data is in general a label map, with label values of low order. In the case of a two class problem as in this paper; bone and no bone, this label map is binary. From an information preserving point of view, the description of a sampled gray-level signal with samples of lower dimensionality, requires a higher sample density. For example, describing an 8-bit signal with binary values would require 8 times the samples to be able to describe the same amount of information. For images, where the signal spectrum is highly biased towards low spatial frequencies, the actual required sample density is lower. However, in CT data, where the size of the structures of interest are comparable to the voxel dimensions (in the slice direction) a higher sampling rate is preferable.

In the CT data used in this paper the voxel dimensions are $0.82 \times 0.82 \times 3 \text{ mm}^3$. We found it necessary to resample the data by a factor of two in the slice direction to adequately be able to describe the segmentation result, i.e. double the number of slices. This was done by upsampling by factor of two in the slice direction followed by smoothing with a $3 \times 3 \times 3$ Gaussian filter kernel, which gives the new voxel dimensions $0.82 \times 0.82 \times 1.5 \text{ mm}^3$. Note that the voxels still

are non cubic. A reason for not expanding the original data more than necessary is that interpolation to higher resolution is a difficult problem in itself [9, 10].

CT data used for the experiments in this paper has voxel dimensions (a, b, c) = (0.82, 0.82, 3), and after resampling (a', b', c') = (0.82, 0.82, 1.5). The ratio between the in-plane and the out-of-plane distance is then, $r = \frac{c'}{a'} = 1.83$

Step 3: Tensor controlled adaptive filtering

The adaptive filtering scheme consists of two main parts: 1) filtering the data with a filter set consisting of six directed quadrature filters producing a description of local structure, and 2) filtering the data with a basis filter set consisting of one low-pass filter and six directed high-pass filters. The final filter result is a combination of the basis filter outputs controlled by the information of the local structure. The affine frequency model discussed in the previous section is applied to all the filters involved.

A tensor **T**, describing local structure, can be obtained by a linear summation of lognormal quadrature filter output magnitudes, $|q_k|$, weighted by predefined tensors \mathbf{M}_k associated with each filter [11]. For the CT data presented in this paper we have experimentally found that a lognormal filter with centre frequency, $\rho_0 = \pi/4$, and relative bandwidth, B = 2, was appropriate for structures inplane, which gives the out-of-plane filter center frequency $r\frac{\pi}{4}$, ((r = 1.83), from above). For details see [11]. It is desirable that the model describing the local neigbourhood varies slower than the signal it is describing. To ensure a slowly varying adaptive filter and to reduce the impact of noise the estimated tensor field, **T**, is relaxed by applying a $7 \times 7 \times 5$ Gaussian filter. The output from this relaxation is denoted \mathbf{T}_r .

The filter set used in the adaptive scheme consists of a low-pass filter and six directed high-pass filters. This basis set can construct filters of a variety of shapes; filters that can smooth along lines, along planes, high-pass filtering across lines or across planes. The basic idea is to compose a filter that fits the local structure of the data,

$$F = F_{lp} + h(\|\hat{\mathbf{T}}_r\|) \sum_k \langle \mathbf{M}_k, \mathbf{C} \rangle F_{hp}^k$$
(4)

where $\hat{\mathbf{T}}_r = \frac{\mathbf{T}_r}{\|\mathbf{T}_r\|_{max}}$ and **C** is the tensor \mathbf{T}_r normalized with the corresponding largest eigenvalue, $\mathbf{C} = \frac{\mathbf{T}_r}{\lambda_1}$. **C** defines the angular behaviour of the adaptive filter. The function, $h(\|\hat{\mathbf{T}}_r\|)$ in Equation 4, controls how much over-all high pass information is allowed through the adaptive filter F. In noisy regions without well defined structures the norm of the tensor $\|\hat{\mathbf{T}}_r\|$ will generally be small compared to areas with well defined structures such as edges. A desired behaviour of the adaptive filter is that when no structures are present, lowpass filtering should be performed in all direction in order to smooth out the noise (isotropic lowpass filtering). Therefore, the function h is designed to be zero for small arguments. The consequence of this will be that the highpass term in Equation 4 is cancelled in regions not having well defined structures and that lowpass filtering is performed instead. The function used is not critical as long as it can be tuned to the described behaviour. The following function was used

$$h(|\hat{\mathbf{T}}_r|) = \frac{|\hat{\mathbf{T}}_r|^{\beta}}{|\hat{\mathbf{T}}_r|^{\beta+\alpha} + \sigma^{\beta}}$$
(5)

with $\alpha = 1$, $\beta = 2$, and $\sigma = 0.05$.

Step 4: Segmentation

This step has been implemented in a very straight forward way. However, despite its simplicity, it was shown to work satisfactorily. The procedure is as follows.

First the adaptively filtered data was thresholded. An example of this step is shown in Figure 2, where the lower left image shows the adaptively filters result, and the lower middle image shows the result of thresholding. Then the thresholded data was labelled into connected components. Isolated islands of voxels inside larger structures were merged into the largest one. The result of this can be seen in the lower right image in Figure 2.

3 Results

Figure 1 displays the results of thresholding a CT data set of a hip joint before and after adaptive filtering. Although the filter was applied in 3D, a single slice is displayed and compared to thresholding at different signal intensity levels. The size of the 6 filters used for the local structure estimation was $9 \times 9 \times 7$, and $15 \times 15 \times 9$ for the 7 basis filters used in adaptive filtering scheme (1 lowpass and 6 directed highpass filters).

The top left image shows the CT scan. This slice shows the hip joint. The top middle image shows the segmentation using a high enough threshold to separate the joint space. Unfortunately important parts of the bone are missing, e.g. the hole in the femoral head (false negatives). The top left image shows segmentation using a low enough threshold to capture most of the important bone. Unfortunately, the result is very noisy and now the femur and the pelvis are connected (false positives). These are the same reasons that make it difficult to capture thin bone structures. Thin bright structures disappear [2] and dark thin structures get filled and this artificially connects different bones (Figure 1 top right). A threshold that separates the joint space inevitably introduces severe artifacts and removes large parts of real bone (false negatives).

The bottom left image shows the result of 3D adaptive filtering of the CT data. The lower right image shows segmentation of the filtered data. The same threshold as in the top right image was used. Note the reduction of the noise level and that the joint space is free from falsely segmented bone structures. Note also that the boundaries of the segmented structures have not moved. This is because smoothing of the data has been performed along the structures and not across them, at the same time extra sharpening is obtained across the structures.



CT images

high threshold

low threshold

Fig. 1. Adaptive filtering of CT data through the hip joint Top: Original gray-scale image (left). The middle image shows the segmentation using a threshold high enough to separate the joint space. The rightmost image shows segmentation using a threshold low enough that is captures all of the important bone. Bottom: The left image shows a gray-scale image as the result after 3D adaptive filtering of the CT data. The lower right image shows segmentation of the filtered data. The same threshold as in the top right image was used.

Figure 2 shows adaptive filtering of a second data set. This data set is displayed differently than in Figure 1. The images presented are slices from the volume orthogonal to the slice direction of the data acquisition. The top left image clearly shows that the voxels dimension are non-cubic. The ratio of the vertical and horizontal resolution is 3/0.82 = 3.7. The top middle image shows the effect of thresholding the original data. It is impossible to find a threshold that separates the femur and the pelvis without removing most of the data. The top right image shows the result of manual segmentation by an orthopaedic surgeon. The lower left image shows the result after resampling the original data by a factor of two in the slice acquisition direction followed by adaptive filtering. The result of thresholding this data is shown by the lower middle image. The two bone parts are now separated and well defined. A connected component labeling and filling of interior regions gives the final segmentation result presented in lower right image.

In Figure 3 the manually segmented femur and the automatically segmented femur are visualized side by side using surface models. The voxel overlap between the two data sets is 98%. In the voxel overlap calculation, the automatically obtained segmentation result was subsampled by a factor of two in the slice direction in order to get the same resolution as the manually segmented data.



Fig. 2. Top: An image from the original CT volume of a hip joint perpendicular to the slice direction (the pelvis is seen in the top part of the images and the femur in the lower part). The middle image shows the result from thresholding the original CT data. The right image shows the result from careful manual segmentation by an orthopaedic surgeon. Bottom: Automatic segmentation using resampling and adaptive filtering to enhance local structure (left image), thresholding of the filtered result (middle), final segmentation by connected component labeling (right).

4 Summary and future work

A three-dimensional adaptive filtering scheme has been implemented for the particular needs of enhancing CT data to improve bone segmentation. It was shown how the frequency characteristic of the filters can be designed to fit the data geometry and compensate for the non-isotropic voxel dimensions normally present in CT data. This also gives well tuned frequency characteristics in all directions of the volume, which is of importance since many of the structures of interest have a size close to the signal spacing. By using adaptive filtering, structures of interest can be emphasized and thus more easily segmented. We have shown that the presented method can automatically segment and separate bones that are close spatially, such as femur and pelvis in the hip joint. For the few cases we have segmented so far, we have obtained around 98% agreement in terms of voxel overlap between automatic and manual segmentation of the data. Future work will be focused on extending our implementation of the linear model in Equation 1, i.e. not only perform scaling of the Fourier domain along the axes when designing filters to get the desired frequency behaviour, but to incorporate shift and shear as well. By introducing subvoxel shifts by modulation in the frequency domain, a resampling to a new grid can be done inside the adaptive filtering scheme.



Fig. 3. Left: Manually segmented femur. Right: Fully automatic segmented femur using adaptive filtering to enhance local structure followed by thresholding and connected component labeling. The voxel overlap between the manually segmented and the automatic segmentation result is 98 %.

References

- 1. G. H. Granlund and H. Knutsson. Signal Processing for Computer Vision. Kluwer Academic Publishers, 1995. ISBN 0-7923-9530-1.
- C-F Westin, A. Bhalerao, H. Knutsson, and R. Kikinis. Using local 3D structure for segmentation of bone from computer tomography images. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 794-800, Puerto Rico, June 1997.
- H. Knutsson, R. Wilson, and G. H. Granlund. Anisotropic non-stationary image estimation and its applications — part I: Restoration of noisy images. *IEEE Trans* on Communications, COM-31(3):388-397, March 1983.
- R. Wilson, H. Knutsson, and G. H. Granlund. Anisotropic non-stationary image estimation and its applications — part II: Predictive image coding. *IEEE Trans* on Communications, COM-31(3), March 1983.
- P. Perona and J. Malik. Scale space and edge diffusion using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7):629-639, July 1990.
- G. Gerig, O. Kübler, R. Kikinis, and F. Jolesz. Non-linear anisotropic filtering of MRI data. *IEEE Transaction on Medical Imaging*, 11(2):221-232, June 1992.
- 7. A. Kumar, A. Tannenbaum, and G. Balas. Optical flow: A curve evolution approach. *IEEE Trans. on Image Processing*, 5(4):598-610, April 1996.
- 8. J.A. Sethian. Curvature and the evolution of fronts. Commun. Math. Phys., 1985.
- 9. J. A. Parker, R. V. Kenyon, and D. E. Troxel. Comparison of interpolation methods for image resampling. *IEEE Transactions on Medical Imaging*, 2(1), 1983.
- 10. C-H Lee. Restoring spline interpolation of CT images. *IEEE Transactions on Medical Imaging*, 2(3), 1983.
- 11. H. Knutsson. Representing local structure using tensors. In The 6th Scandinavian Conference on Image Analysis, pages 244-251, Oulu, Finland, June 1989.